Department of Molecular Biology
Faculty of Sciences
Autonomous University of Madrid (UAM),
Madrid, Spain

Characterization of a native plasmid from *Bacillus subtilis* with special focus on its regulatory circuit for conjugation

Doctoral Thesis

Praveen Kumar Singh

Madrid, Spain, 2014

Thesis presented by Praveen Kumar Singh to obtain Doctor of Philosophy in science degree at the Autonomous University of Madrid (UAM), Madrid, Spain.

This Work has been done at Centro de Biologia Molecular "Severo Ochoa" (CSIC-UAM) under the supervision of Dr. Wilhelmus J. J. Meijer with the financial support from Consejo Superior de Investigaciones Científicas.

Dedicated to Papa and Mummy

Acknowledgements

Finally I have reached the stage where I want to thank everyone who directly or indirectly helped me during my Ph.D.. First, I feel my sincere gratitude to my supervisor Dr. Wilfried whose continuous support and motivation gave me so much strength to finish my thesis with a good note. Wilfried was not only my supervisor but he was also my friend. When I came to Spain from India, I had no idea about language and neither about European culture, but Wilfried was always there to help me. Working under his supervision and with his enthusiasm at every step has made me a better researcher. during my stay in the CBCB, Newcastle University (UK), I have learned many important techniques from Dr. Ling Wu. I would like to thank Proff Jeff Errington who gave me space in his lab to work. All the members of CBCB, Newcastle have been integral part of my work and I thank all of them. My second short stay was very fruitful and that happened only because of kind help from the members of CBC, Swammerdam Institute of Life Sciences, University of Amsterdam (The Netherlands) and specially Dr. Leendert Hamoen. He always had time to answer my questions and to see my results. I have been provided full independency to use the most advanced microscope present in his lab and thus I learned a lot in microscopy techniques. Special thanks to Riya who helped me how to make good slides for live microscopy.

I would like to thank to all my friends who were very special to me during my PhD study and also my stay in Madrid, Spain. I am starting with Tribhuwan who was the only my friend when I came to Madrid. He provided everything to me and our friendship has become stronger after living together for five years in the same house. We shared all our thoughts, all our feelings and even we sometimes helped each other with our experiments. Those days were my best days and I really enjoyed my stay with Tribhuwan.

Having a good friend in lab has always benefitted me in many ways. Gayetri has played very important role during my PhD studies. Be it discussion or designing and performing an experiment, Gayetri had always been there. I would have not published so many articles in such a limited time, if she had not done crucial experiment after reviewer's comments. I cannot forget cloning of *rok* gene how she had done when I needed it urgently. We also shared our good and bad times during my stay in Madrid.

Even being far from me and living in different countries my old friends have always been there whenever I needed. Jhauwa, Shukla Ji, Amirchand, Goutam Tanti,

Bhaskarwa, Imran, Raghav and Amar baba, all of you became part of my memory, which will always be there for my whole life. How can I forget our hangout over google talk.

My best Spanish friend, Andres had taught me about Spanish culture and I started speaking Spanish only after I met Andres. Muchas gracias Andres para todos.

During my stay in Spain, I made new friends Valencio and Hena. Valencio always cooked fish during weekends and I always enjoyed eating it. Thanks to both of you for being there and helping me. My housemates and friends, Mayank and Acharya, made my last days very pleasant. I would also like to thank my labmates Cesar, Jorge, Esther and Lucia.

I would have never achieved this without sacrifice of my family. They always were behind me and promoting me even during my bad times. My Papa always enjoyed discussing my findings with his friends. My Mummy is sweetest member of my family. She is always worried if I eat properly or not. Praduman bhaiya, who was always ready to listen my scientific findings and his positive attitude towards life has given me strength to go ahead and work hard. My Uncles, Samar Cha and Sher Cha, were also like my Papa and they always discuss my achievements to their friends and really felt proud of me. Last year, Neha joined my family as my wife. I feel really happy after marrying Neha. She enjoyed my scientific discussion and always ready to listen me. She had to stay far from me during my last days but she has never complained about it. Thanks Neha for understanding me.

I would like to thank my collaborators Ramon, Ricardo and Ramos for helping me with RNAseq. I have been lucky in getting fellowship and would like to acknowledge Consejo Superior de Investigaciones Científicas (CSIC) for helping me economically.

I might have missed mentioning many names that had played an important role during my Ph.D. thesis. I would like to thank to all of them for their valuable help.

Summary

Spreading of antibiotic resistance among pathogenic bacteria is a major problem for human health as well as in veterinarian world. Bacteria can exchange antibiotic resistance genes by different processes, collectively called Horizontal Gene Transfer (HGT). One of the processes of HGT is conjugation in which genetic material (DNA) is transferred from a donor cell to a recipient cell. Conjugation, which is widespread among both Gram-negative and Gram-positive bacteria, is mediated by conjugation system present on plasmids and/or by integrative conjugative elements (ICEs). Plasmids are autonomously replicating units present in many bacteria besides their chromosome. Conjugation systems present on various plasmids from Gram-negative bacteria have been studied in considerable detail. Much less is known, however, about conjugation systems present on plasmids from Gram-positive bacteria. A major goal of my thesis studies was to increase our understanding about conjugation systems and its regulation in Gram-positive bacteria. We chose the native plasmid pLS20 from *Bacillus subtilis* for several reasons. First; *B. subtilis*, itself a GRAS organism, is related to pathogens or fastidious bacteria, 2nd, is one of best studied organisms, 3rd, its natural competence facilitates genetic manipulation, and 4th, *B. subtilis* is gut commensal.

The first part of this thesis describes an *in-silico* analysis of pLS20cat, which we have sequenced to completeness. pLS20cat has a size of 65,774 bps. It contains almost 100 ORFs, which we have grouped into five different modules. One of the modules concerns a putative conjugation operon, which flanks a *rap-phr* cassette. Many *rap-phr* cassettes have been identified on the genome of *B. subtilis* and other bacilli. In addition, they are present on several rolling-circle and theta replicating plasmids of Bacilli. Most of the rap-phr cassettes analyzed have been shown to affect differentiation processes like sporulation, competence and biofilm formation. A detailed analysis of the *rap-phr* cassette of pLS20cat is described in the third part of this thesis.

We found that the presence of pLS20cat strongly inhibits transformation efficiencies of its host. The second part of this thesis describes the identification and functional analysis of the pLS20cat-located gene responsible for competence inhibition. Thus, we found that ORF64 is responsible for competence inhibition and we named the gene *rok*$_{LS20}$. We also unraveled the mechanism of how Rok$_{LS20}$ inhibits competence. Rok$_{LS20}$ is a DNA binding protein that preferentially binds to the *comK* promoter thereby repressing expression of the key transcriptional activator of the competence genes, ComK. A functional Rok$_{LS20}$-GFP fusion protein co-localized with the bacterial nucleoid in an irregular pattern similar to that

described for the Rok_{Bs168} homologue that is encoded by the host. Screening the sequences present in available databases resulted in the identification of several additional *rok* paralogues. Based on their homology, 20 *rok* genes were divided into two groups.

The third part of the thesis describes the regulatory circuit of pLS20cat conjugation. Our results show that the conjugation genes of pLS20cat are not induced by recipient-produced pheromones, demonstrating that the conjugation genes of pLS20cat are fundamentally different regulated to those present on the Gram-positive enterococcal plasmids pAD1 and pCF10. Our results show that conjugation is in a default "OFF" state. The pLS20cat-encoded Rco_{LS20} protein, which belongs to the Xre family of transcriptional repressors, is the main transcriptional regulator that keeps conjugation in the "OFF" state. Conjugation is activated by Rap_{LS20}, which is an anti-repressor and hence relieves Rco_{LS20}-mediated repression. Finally, the phr_{LS20} encodes for the signaling peptide Phr^*_{LS20} that is responsible for inhibiting the activity of Rap_{LS20} and decide the timing for conjugation.

Abbreviations

Abbreviations

DNA	Deoxyribonucleic Acid
RNA	Ribonucleic Acid
HGT	Horizontal Gene Transfer
ICE	Integrative and Conjugative Elements
cat	Chloramphenicol Acetyltransferase
IPTG	Isopropyl β-D-1thiogalactopyranoside
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside
T4CP	Type IV coupling protein
Mfp	mating pair formation
QS	Quorum sensing
Inc	Incompatibility
LB	Luria Broth
MM	Minimal Medium
PAGE	Polyacrylamide gel electrophoresis
EtBr	Ethidium Bromide
SDS	Sodium dodecyl sulphate
nt	nucleotide
bp	Base pair
C-ter	C-terminal
N-ter	N-terminal
TAE	Tris-acetate EDTA
TBE	Tri-borate EDTA
BSA	Bovine serum albumin

Contents:

1. Introduction

1.1 General Introduction

Bacteria are amazingly adaptable little critters. They are constantly evolving to become better suited to their environment and even they quickly develop resistance against antibiotics. The first time antibiotic resistance in pathogens emerged as a major problem was during the 1950s when Japanese hospitals encountered *Shigella* dysentery outbreaks that resisted to treatment with usual antibiotics. Intensive clinical and genetics investigations performed by Japanese scientists during that period resulted in a concept of episome-mediated transfer (a type of Horizontal Gene Transfer, HGT) of the drug resistance in the Enterobacteriaceae (1, 2). This research also resulted in the eventual use of antibiotic resistance markers for genetic manipulations and was, therefore, one of the foundations for the development of genetic engineering. The genetic engineering tools developed were used to study the molecular mechanisms of antibiotic resistance. Increasing numbers of tools like DNA hybridization, sequencing, and PCR were used for the analysis of gene exchange processes in natural settings (3–5). The important role HGT plays in the evolution of bacterial species especially highlighted during the last decade with the advent of high-throughput sequencing. Comparative genomics analyses revealed that substantial parts of bacterial genomes consists of auxiliary genes acquired by HGT (6). These genes may confer adaptive advantages under specific environmental conditions, for instance the presence of antimicrobials, xenobiotics, heavy metals and other compounds. Horizontally acquired genes may also allow the colonization of new ecological niches governed by biotic factors such as symbiotic and pathogenic relationships (7). In summary, HGT has been proven to be very important in the field of medical microbiology, biotechnological studies and evolutionary biology.

HGT can occur by three main mechanisms: 1) transformation, 2) transduction and 3) conjugation. Conjugation has the broadest host range (BHR) of transfer and thus has maximum potential to contribute towards HGT. Conjugation is mediated mainly by plasmids and integrative conjugative elements (ICEs). Plasmid conjugation is the principal decisive factor for BHR potential of conjugation. BHR plasmids are able to transfer across different phyla and sometimes even can cross the prokaryote/eukaryote barrier. Thus, plasmids can be advantageous to its hosts but they are also responsible for antibiotic resistance determinants or disease spreading. It is crucial to study the basic mechanisms involved in plasmid transfer and regulatory processes that control conjugation.

The process of conjugation and its transcriptional regulation has been studied in considerable detail for various plasmids present in Gram-negative bacteria (for review see, 8–11). However, comparatively little is known about conjugation systems on plasmids from Gram-positive bacteria, many of them industrially and medically important organisms, although interest in this field is increasing (for general review see, 10–12). For a better understanding of plasmid conjugation, we chose natural plasmid pLS20 first identified in *B. subtilis* natto strain IFO3335 (13). Studies carried out in this thesis started with sequencing and annotating plasmid pLS20cat. pLS20cat is a derivative of pLS20 labeled with a chloramphenicol resistance gene (14). The main focus was on understanding the regulatory circuit of pLS20 conjugation. We have characterized the principal factors regulating pLS20 conjugation, which are the main repressor of conjugation (Rco), quorum-sensing peptide (Phr) and an antirepressor that belongs to a large family of proteins, Rap, which are mainly encoded by bacilli. We also found that the presence of pLS20cat inhibits competence of its host. We have identified the pLS20 located gene responsible for this. The biological significance of this gene with respect to evolution has also been discussed. We used advanced techniques like RNAseq to analyze pLS20 transcriptome and fluorescent microscopy the subcellular localization of proteins in living cells. In brief, this is the first in-depth study on the regulation of plasmid conjugation from a fertility plasmid of *Bacillus subtilis.* The results presented in this thesis have advanced our knowledge on regulation of plasmid conjugation in *B. subtilis* and in Gram-positive bacteria. In general, in the following sections of the introduction the current knowledge of the different type of mechanisms responsible for HGT, detailed mechanism of plasmid conjugation and types of plasmids, followed by the regulatory systems a few prototypes of conjugation systems will be described. Brief introduction of pLS20 has been highlighted in the end.

1.2 Horizontal Gene Transfer

Horizontal (lateral) gene transfer, the transfer of genes between different species, is an evolutionary phenomenon. HGT is recognized as one of the most troublesome aspects of antibiotic resistance because it can rapidly move resistance genes among pathogenic, commensal, and environmental bacteria. Recent analyses of genomics show that HGT is a potent force in prokaryotes. The percentage of genes in some species obtained by HGT has been calculated (15). A summary of these studies is given in Table 1 (modified from, 15), which demonstrate that many (pathogenic) bacteria contain significant percentage of genes in its genome by HGT.

HGT occurs by three main mechanisms: 1) transformation, 2) transduction and 3) conjugation (6, 16, 17). A description of each of these three mechanisms is described below. Another, recently described mechanism of HGT called phage-related chromosomal islands (PCRI), is not described here (18).

Table 1: Species, Disease caused and Percentage of HGT

Species	Disease caused	Percentage* HGT
Proteobacteria		
Escherichia coli		9.62
Haemophilus influenza	Pneumonia	6.19
Helicobacter pylori 26695	Ulcer	6.41
Helicobacter pylori j99	Ulcer	5.81
Rickettsia prowazekii	Typhus	3.62
Gram-positive Bacteria		
Bacillus subtilis		14.47
Mycoplasma genitalium	Urethritis	14.47
Mycoplasma pneumoniae	Pneumonia	5.93
Mycobacterium tuberculosis	Tuberculosis	5.01
Spirochaete		
Borrelia burgdorferi	Lyme disease	1.56
Treponema pallidum	Syphilis	8.32
Chlamydiae		
Chlamydia trachomatis	Trachoma, epidymitis	4.32
Chlamydia pneumoniae	Pneumonia, bronchitis	5.7
Aquiflex aeolicus		4.84
Deinococcus radiodurans		3.92
Synechocystis PCC6803		7.50
Thermatoga maritima		11.63

*The percentage HGT was calculated by dividing the number of proposed transferred genes by the sum of nonproposed ones, excluding genes smaller than 300 bp, proposed transferred genes.

1.2.1 Transformation

Transformation is the active uptake of exogenous DNA by competent cells usually followed by genomic integration mediated by homologous recombination. Griffith performed the first transformation experiment in *Streptococcus pneumoniae* in 1928, and Avery and coworkers confirmed in 1944 that DNA is the genetic material that is transformed. Since then transformation has been recognized as a powerful mechanism of HGT in natural bacterial populations. Naturally transformable species have been detected in almost all the major taxonomic groups of Eubacteria as well as in the Archaea (19). Competence genes are usually encoded by the bacterial chromosome and often dispersed across regions (19).

For natural transformation to occur, DNA has to be released from donor cells and dispersed or maintained in the environment until being encountered by potential competent cell. Studies demonstrated that DNA is a main component of the extracellular

matrix of colonies of probably most bacterial species and is present at high concentrations in diverse prokaryotic habitats and persist for considerable time (20). Supporting to this view, a recent study provides evidence that the Gram-positive *Bacillus subtilis* cells secrete DNA actively in its environment during competence development (21). Development of competence involves the regulated expression of genes encoding for proteins that constitute the sophisticated DNA uptake system. This DNA uptake machinery is related to type IV pili and type II secretion systems (19, 22). Figure 1 summarizes transformation steps as 1) absorbing foreign ssDNA by competent cell and 2) recombination of exogenous DNA with host chromosome resulted in new genetically transformed cells.

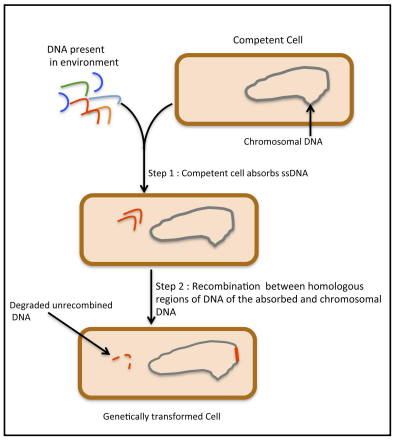

Figure 1: Basic steps of natural transformation in bacteria

In this thesis we described that pLS20, a native plasmid from *B. subtilis,* inhibits competence development of its host. The results obtained provide possible clues for differences in competence levels that are developed by wild type *Bacillus* strains.

1.2.2 Transduction

Phages play an important role in gene transfer in the environment because of their wide occurrence. Some phages are able to "mobilize" bacterial genes by encapsidating the corresponding DNA into the phage head and transfer into another cell by injection. For a long time, phages were assumed to be species specific and rarely crossed boundaries of species. However, the discovery of broad-host-range (BHR) generalized transducing phages (23, 24) changed this view completely. There are few examples where transduction can even occur between bacteria belonging to different classes (25) and the range of organisms into which a phage can inject its DNA can be wider than the infective host range (25, 26).

The first step of phage infection consists of attachment of the phage to cognate receptor molecules on the surface of the bacterial cell. Phages attach to their host receptors via specific molecules, known as antireceptors. The specificity with which phage antireceptors recognize the bacterial receptors confines the host-range of a phage. Phage antireceptors are often flexible and able to recognize several bacterial surface molecules (27, 28). There are even cases known where phages contain multiple antireceptors genes for different receptors, and hence displaying more than one

antireceptor in the same phage (29, 30). Thus, phages evolved different machineries to invade its host and at least some of them are able to mobilize bacterial genes and transfer them to other bacterium.

There are two types of transduction: generalized and specialized (31, 32). Generalized transduction can transfer any host gene and it occurs when phage packaging accidentally incorporates bacterial DNA instead of phage DNA (32–34). Specialized transduction is due to faulty excision of the prophage from the bacterial chromosome. The schematic diagram given in Figure 2 explains the mechanism of transduction in different steps as:

Step (1) Phage attaches to its host surface and injects its DNA,

Step (2) Phage enzyme degrades host DNA.

Step (3) Phage progeny is synthesized inside the host cell that has incorporated phage DNA and, mistakenly, some host DNA.

Step (4) Transducing phage (with donor DNA) injects its DNA into a new host.

Step (5) Injected DNA incorporated into the chromosome of the infected cell by recombination.

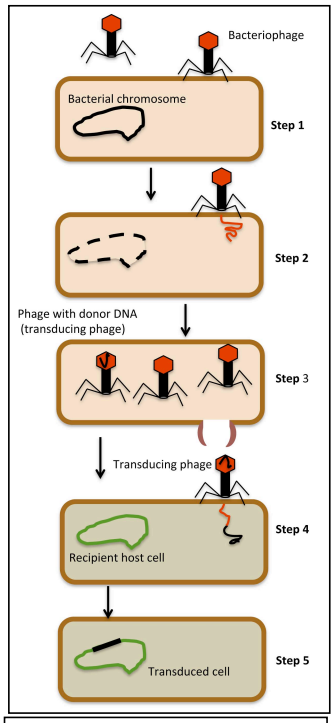

Figure 2. Basic steps of HGT mediated by transduction

1.2.3 Conjugation

Bacterial conjugation is a highly specific process whereby DNA is transferred from donor to recipient cells by a specialized multiprotein complex, termed the conjugation apparatus (35) .Contrary to competence and transduction, conjugation requires direct contact between the donor and recipient cells. A single donor cell can convert a population of recipient cells to donor cell status via the process of epidemic spread, which is a feature of conjugation alone. The effectiveness of conjugation in disseminating traits such as antibiotic resistance is remarkable, with resistance to newly available antibiotics appearing within months of their introduction (36). Conjugation has the broadest host range of transfer compared to other bacterial gene exchange mechanism (37, 38). Conjugation is widespread among both Gram-negative and Gram-positive bacteria and it is mediated by plasmids and/or by integrative conjugative elements (ICEs).

The basic mechanism of conjugation for all transmissible plasmids/elements is similar. A schematic overview of conjugation is given in Figure 3. The donor cell

produces a mating pair formation (Mpf) system that enables intimate physical contact with a recipient cell through a membrane-spanning protein complex and a surface-exposed sex pilus (step 1), at the same time relaxase, which is a phosphodiesterase, cleaves substrate DNA in a site- and strand-specific manner within the origin of transfer (*oriT*) generating ssDNA and remains covalently joined to the 5'DNA end, resulting in a nucleoprotein complex called the relaxasome (39)(40) (step 2), The 'coupling protein' (T4CP) links the relaxasome to the secretion pore formed by the MpF system, through which the ssDNA passes on its way to the recipient cell (41) (step 3), and conversion of received ssDNA into dsDNA in the recipient cell (step 4). New donor cell is now ready to transfer its plasmid to potential recipient same as mentioned earlier. Mpf/CP conjugation systems belong to the family of type IV secretion systems (42). Much less is known about how DNA crosses the recipient cell membranes.

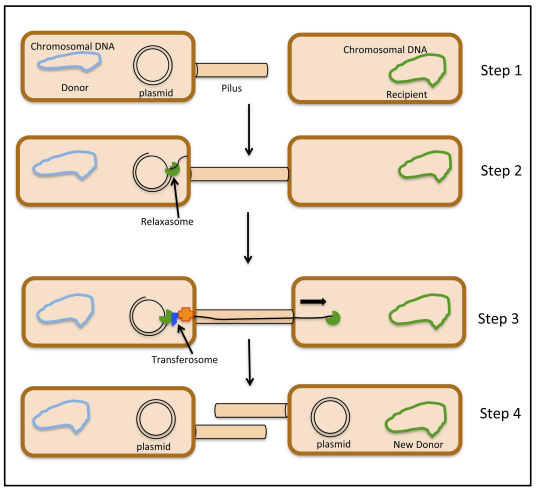

Figure 3. Basic steps involved in plasmid-mediated conjugation.

In Gram-positive bacteria, some conjugative elements produce surface-presented protein factors that induce cell aggregation (43). Very little is known about conjugation systems on plasmids from Gram-positive bacteria compared to that of Gram-negative bacteria (for general review see, 47).

1.2.3.1 Conjugative Plasmids (self-transmissible)

Plasmids are key players of horizontal gene transfer and important tools in genetic engineering. They code for proteins involved in detoxication, virulence, ecological interactions and antibiotic resistance. Hence, to understand the evolution of these important bacterial traits, often involved in human health or well-being, it is essential to understand plasmid conjugation (for review see, 9). C

Conjugative plasmids encode the complete protein machinery for and all other proteins required for its conjugal transfer. Hence these plasmids are also called self-transmissible. The ability to transmit horizontally allows the plasmid to disseminate throughout different hosts present within a certain niche. The study of conjugative plasmids was stimulated by the explosive emergence and spread of antibiotic resistance

that followed shortly after entering into the era of antibiotic treatment for bacterial infections. Thus, the majority of conjugative plasmids were studied because, in one way or another, they affected the well being of humans. Till date, many conjugative plasmids have been discovered from both Gram-negative and Gram-positive bacteria. A list of various conjugative plasmids from Gram-positive bacteria is given in Table 2 (modified from, 47).

Table 2: Conjugative plasmids, size, original Host and resistance to different antibiotics

Plasmid	Size (kbp)	Original host	Resistance
pIP501	30.2	*Stahylococcus agalactiae*	Cm, MLS
pAMß1	26.5	*Enterococcus faecalis*	MLS
pRE25	50.2	*Enterococcus faecalis*	Cm, MLS
Tn916	18.0	*Enterococcus faecalis*	Tc
Tn1545	25	*Stahylococcus pneumoniae*	Em, Km, Tc
pSK41	46.4	*Stahylococcus aureus*	Bm, Gm, Km, Nm, Tm
pGO1	52.0	*Stahylococcus aureus*	Bm, Gm, Km, Nm, Tm, Tmp
pMRC01	60.2	*Lactobacillus lactis*	None
pAD1	59.3	*Enterococcus faecalis*	None
pCF10	65	*Enterococcus faecalis*	Tc
pPD1	56	*Enterococcus faecalis*	None
pXO16	200	*Bacillus thuringiensis*	Unknown
pRS01	48.4	*Lactobacillus lactis*	Unknown
pLS20	65.7	*Bacillus subtilis natto*	Unknown

Drug resistance abbreviations: Cm- chloramphenicol, MLS- Macrolide-lincosamide-streptogramin B, Tc- tetracyline, Em-erythromycin, Bm- bleomycin, Km- kanamycin, Nm- neomycin, Tm-tobramycin, Tmp- trimethoprim, Gm- gentamicin

1.2.3.2 Mobilizable plasmids

This type of plasmid carries the genetic information necessary for relaxasome formation and processing, but lacks the functions required for mating pair formation. These plasmids can be transferred horizontally to other bacteria if a membrane-associated mating pair formation (MPF) complex is provided by another self-transmissible element present in the same cell. Mobilizable plasmids have been classified according to their relaxases (45). Table 3 lists some mobilizable plasmids from each relaxase superfamily.

Mobilizable plasmids have a large impact in horizontal gene transfer in nature, including the spread of antibiotic resistance. However, mobilizable plasmids have been studied in less detail than to conjugative plasmids (for more detail see review, 49)

Table 3: Some mobilizable plasmids, their hosts and superfamily			
Plasmid	**Size (kbp)**	**Bacterial Source**	**Superfamily**
RSF1010	8.684	*Escherichia coli*	MOB_Q
pIE1130	10.687	*Environmental sample*	MOB_Q
pAB6	5.597	*Neisseria meningitidis*	MOB_Q
pAsal2	5.424	*Aeromonas salmonicida*	MOB_P
pSK639	8.013	*Staphylococcus epidermidis*	MOB_P
pS194	4.397	*Staphylococcus aureus*	MOB_P
ColE1	6.650	*Escherichia coli*	MOB_{HEN}
pC	5.269	*Salmonella enteritidis*	MOB_{HEN}
pLG13	6.293	*Escherichia coli*	MOB_{HEN}
pMV158	5.536	*Streptococcus agalactiae*	pMV158
pI4	14.000	*Bacillus coagulans*	pMV158
pBC16	4.630	*Bacillus cereus*	pMV158
CloDF13	9.957	*Enterobacter cloacae*	CloDF13

1.2.3.3 Integrative and Conjugative Elements (ICEs)

ICEs are a diverse group of genetic elements that encode the functions required to integrate into their bacterial host chromosome and to transfer themselves between cells by conjugation. As illustrated in Figure 4, upon induction, ICEs excise themselves from the chromosome by site-specific recombination and form a circular intermediate (steps 1 and 2). Next, ssDNA generated from resulting circular intermediate is transferred through a mating pore encoded by ICEs (step 3) and integrate by recombination between a specific site of this transferred DNA and a site in the genome of their host (step 4) (46–48).

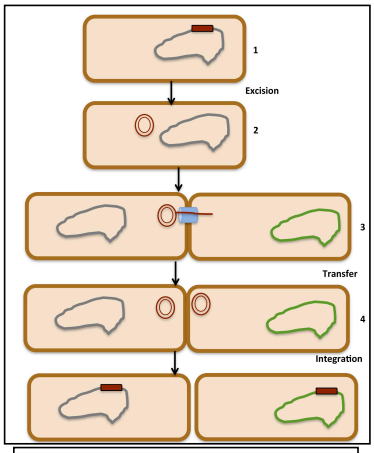

Figure 4: Basic steps involved in transfer of ICEs

The ICEs include conjugative transposons and conjugative transposon-like genomic islands, as well as numerous unclassified mobile genetic elements (MGEs). Genome sequence analyses suggest that ICEs are widespread in bacteria and they probably contribute importantly to horizontal gene transfer. Like other MGEs, host range for ICEs also varies like other MGEs (49).

1.3 Regulation of Conjugation Machinery

Conjugation systems are present on many plasmids present in Gram-positive or Gram-negative bacteria. Conjugative elements have evolved regulatory systems to minimize the metabolic and phenotypic load that is imposed on the host by the synthesis of the conjugative transfer apparatus while optimizing the adaptive advantages of self-transmission. Regulatory circuits of conjugation systems present on severl plasmids replicating in Gram-negative bacteria has been studied in considerable detail. Less is known about regulatory circuits present on ICEs and especially on plasmids of Gram-positive bacteria.

Conjugative systems present on various plasmids from Gram-negative bacteria and a few plasmids replicating in Gram-positive bacteria (chiefly enterococci) have been studied in considerable detail (for review see, 12, 41, 50) . Conjugative systems are also present on transferable elements that are integrated in bacterial genomes known as integrative and conjugative elements (ICEs) and the regulatory system of some ICEs have been studied in considerable detail (51, 52). A shared feature of the regulatory circuits of conjugative systems is that transcriptional and post-transcriptional mechanisms are used to ensure that expression of the transfer genes in not an undue burden on the host cell. In general, conjugative systems are in a default "OFF" state and their transfer is induced in response to environmental and physiological signlas as well as by the presence of recipient cells. Detailed knowledge about the mechanism of regulation is available for only a few transfer systems and the last review dates from several years ago (53, 54). Here summary of a few regulatory circuits known from both Gram-positive and Gram-negative bacteria is given. The different regulatory mechanisms have been classified based on the conjugation-inducing signal, being (i) a peptide, (ii) stress, or (iii) other.

1.3.1 Peptide-mediated regulation of conjugation

1.3.1.1 Quorum sensing regulated conjugation

Quorum sensing (QS) is a common way by which bacteria communicate with one another using small and diffusible chemical signaling molecules. When the concentration of a signaling molecule reaches a certain ''quorum'', bacteria respond by altering its gene expression profile at a (sub) population-wide scale (for review see, 55, 56). In essence therefore, QS serves as a simple indicator of population density. In most cases, QS serves to activate a set of genes or a differentiation pathway when a cell population reaches a certain density. However, QS serves to repress the genes

responsible for the conjugative transfer of the ICE*Bs1*, which is present in several *B. subtilis* strains.

ICE*Bs1*

An integrative and conjugative element named ICE*Bs1* is present in the genome of the *B. subtilis* 168 strain used in various laboratories, but is not present in other 168 derived strains as PY79. ICE*Bs1* conjugation is kept in the default "OFF" state by a master regulator that represses expression of the conjugation genes called as ImmR. ICE*Bs1* contains a *rap-phr* cassette (*rapI-phrI*), which is located near the right end of the element (see Fig. 5A). RapI can activate the conjugation genes by relieving ImmR-mediated repression of the conjugation genes, and the mature signaling peptide derived from PhrI regulates activity of RapI. Efficient conjugation of ICE*Bs1* occurs on solid medium and when cells are in the stationary phase. Concordantly, *rapI* of ICE*Bs1* appears to be repressed during the exponential growth phase by the global regulator AbrB (57).

In summary, as shown in Figure 5A, ImmR inhibits the expression of excisionase (xis) and additional downstream genes necessary for ICE*Bs1* excision and transfer. Expression of the *rapI-phrI* cassette during stationary phase results in accumulation of intracellular RapI that results in relief of ImmR-mediated repression of the conjugation genes. The activity of RapI is regulated by PhrI whose concentration depends upon the ratio of donor and recipient cells (47, 58, 59).

Consequently, when most or all cells within a population contain the conjugative element and extracellular levels of the signaling peptides are high, conjugation will be inhibited. On the contrary, conjugation will be activated when many recipient cells are present within the population. Thus, in this case QS is used to arrange that conjugation genes become activated only when recipient cells are potentially present.

1.3.1.2 Pheromone induced conjugation

In the case of the conjugation systems present on ICE*Bs1*, accumulation of the signaling peptide results in repression of the conjugation genes. Peptides can regulate conjugation also in a fundamentally different way, in the sense that they activate conjugation rather than inhibiting it. Recipient-encoded peptides that induce are often referred to as pheromones. One example is the conjugation system present on the pTI plasmid of the Gram-negative *Agrobacterium tumefaciens* responsible for transfer of the plasmid into plant cells, which can lead to induction of tumors. Activation of pTI conjugation requires two signaling peptides, one produced by the plant and the other by the donor cell (60). pTI conjugation concerns transfer of bacterial DNA into eukaryotic cells and will not be discussed here further.

Pheromone-induced conjugation also occurs for the enterococcal plasmids pPD1, pAD1 and pCF10 which are representatives of a large family of conjugative plasmids collectively known as pheromone-responsive plasmids (for review see 61–64). The general regulatory systems of these plasmids have similar basics, although differences between plasmids exist. Here, we summarize as a prototype pheromone induced regulation of plasmid pCF10.

pCF10

This pheromone responsive enterococcal plasmid has a size of 67,673bp. The pCF10 pheromone–sensing system involves two signaling molecules. One of them is peptide cCF10, whose gene is located on the bacterial chromosome. The other signaling peptide, iCF10, is encoded by plasmid pCF10. During the un-induced state, pCF10 containing cells secrete a mixture of both peptides. However, the levels of secreted iCF10 is about 80-fold higher than that of cCF10 because the secretion of the cCF10 peptide is reduced by a mechanism that involves the pCF10-encoded PrgY protein, (65). Upon binding to iCF10, PrgX adopts a repressive complex that binds to the main conjugation promoter P_Q (66). However, promoter P_Q is not fully repressed under these conditions allowing synthesis of the small transcript Qs, which aborts at terminator IRS1 and does not lead to production of conjugation proteins (see Fig. 5B). Qs transcript encode for the iCF10 peptide. At the same time, transcripts derived from the P_x promoter generate sufficient amounts of PrgX to keep P_Q activity below the threshold required for expression of conjugation genes. Besides this main regulatory pathway, regulation of pCF10 conjugation genes also involves a countertranscript-mediated mechanism (67). A change in the ratio of the two peptides can occur either by the presence of recipients, -that secrete cCF10 but no iCF10-, or by rapid diffusion of iCF10, which occurs when donor cells enter the bloodstream. In the first case, the cCF10-PrgX complexes formed cause an allosteric conformational change and result in de-repression of promoter P_Q (66). Consequently, transcripts starting from the P_Q promoter extend past IRS1 terminator and downstream conjugation genes become transcribed resulting in conjugation (68). At the same time, the activity of the P_X promoter drops significantly, and PrgX protein levels also decrease modestly. In the second case, i.e. in bloodstream-related infections, a host factor, probably an albumin/lipid complex, selectively sequesters or degrades iCF10, leading to induction (by endogenous cCF10) of conjugation proteins.

Derepression of P_Q promoter during induced state increases transcription of prgQ operon that leads to overproduction of iCF10. As a result, this favors formation of

the iCF10-PrgX repressor-complex which again turns off expression of the conjugation genes. Thus, the synthesis of iCF10, its processing time, export and reimport decide the short-time window of high-level prgQ operon expression and conjugation of pCF10 (69).

1.3.2 Stress induced activation of conjugation

Different mechanisms have been described above how conjugation genes are activated directly or indirectly by peptides that are an indication of the presence/absence of recipient cells. Conjugation genes may also be activated for very different reasons, being adverse conditions like UV irradiation, nutrient deprivation or other environmental factors that causes stress to the donor cell. It is worth mentioning that adverse conditions are also a major cause for temperate phages to activate their lytic pathway resulting in the liberation of phage progeny in the environment. Therefore, these responses can be regarded as a way to escape a host whose future integrity is uncertain and replace it for another one. As outlined below, this strategy is adopted by ICEs and by some plasmids if we consider that adverse conditions can upregulate conjugation genes.

Stress-induced conjugation of ICEs

Stress-induced conjugation has been described for several ICEs (for review see, 71). DNA damaging agents or other insults provoke the cell to produce SOS response. Amongst others, this results in RecA-stimulated autocleavage of repressors of some ICEs causing them to become activated. This same mechanism is also responsible for activating the lytic pathway of temperate phages. A special case is ICE*Bs1* of *B. subtilis*. Above, it is explained that conjugation of ICE*Bs1* is activated by a QS mechanism involving a *rap-phr* cassette. Interestingly, the ICE*Bs1* conjugation genes are also induced when cells are insulted leading to induction of the SOS response (47). However, inactivation of the ICE*Bs1* repressor ImmR is not a consequence of RecA-stimulated autoproteolyis of ImmR, but requires the ICE*Bs1* endcoded metallo-peptidase ImmA (58) (see Fig. 5A). In other cases, sublethal concentrations of antibiotics can also induce conjugation of ICEs; for example ciprofloxacin may promote the spread of SXT and tetracycline enables transcription of the *tra* genes required for transfer of Tn916 (71, 72).

pSLT

pSLT is an F-like conjugative plasmid from *Salmonella enterica*. The control of conjugal transfer in the F-like plasmids usually relies on regulatory elements encoded on the plasmid itself and has been well described (see review, 73). In addition to the

above mentioned mechanism, in pSLT plasmid it has been shown that host-encoded functions are also involved in the circuitry that governs mating as illustrated in Figure 5C (74). Under aerobic conditions, succinate dehydrogenase (SdhABCD) acts as a repressor of *traJ* transcription (in an unknown manner) keeping conjugation in the "OFF" state. Under conditions of microaerobiosis, a signal transduction system responsive to the oxygen level, ArcAB, induces the conjugation genes of pSLT (75). Binding of the transcription factor ArcA upstream of the main *tra* promoter (P_{traY}) activates expression of the *tra* genes (76). Additionally, ArcA also activates *traJ* transcription indirectly by inhibiting Sdh synthesis. TraJ initiates conjugation by activating TraY, but the mechanism of activation is not clear. Another environmental condition that influences pSLT transfer is the nutrition level. Leucine-responsive regulatory protein (Lrp) co-ordinates cell metabolism in response to the availability of nutrients in the external environments. In case of nutrient deprivation, Lrp autogenously enhance its own transcription and also induces *traJ* transcription and in turn pSLT conjugation (77). Lrp has also been shown to play an important role in the expression of some pilus genes and in pilus phase variation in some *E. coli* strains (for review see, 78). The assessment of metabolic conditions via Lrp control might provide and obvious advantage as pilus building as well as DNA transfer are energy consuming processes.

1.3.3 Other systems

Above, we have described various mechanisms, which ensure that conjugation systems are expressed only when recipient cells are potentially present, or when the integrity of donor cell is compromised. However, in some cases the decisive factors responsible for activating the conjugation genes are not (fully) clear. Below, we mention a few of these cases.

pIP501

The 30.6 kb plasmid pIP501, originally identified in the Gram-positive bacterium *Streptococcus agalactiae* (79), exhibits a very broad host range. Besides being able to replicate in Gram positive streptococci, lactobacilli, enterococci, lactococci, staphylococci, bacilli, listeria and clostridia, it can also replicate in the Gram negative *E. coli* (80). The first gene of the putative transfer operon (*tra*) is named *traA* (see Fig 5D). The TraA protein encoded by pIP501 is a DNA relaxase. The nicking and closing activities of TraA are required to initiate the sysnthesis of ssDNA that is transferred into

the recipient cell. Thus, it introduces a site- and strand-specific nick at the origin of transfer (*ori*T) allowing one strand to be displaced as a consequence of rolling circle type of replication by extending replication at the OH group of the 3´-end at the nick site. Interestingly, TraA also functions as a repressor of the P_{tra} promoter (80). The P_{tra} promoter, which overlaps partially with *oriT*, is located upstream of *traA*, and controls expression of *traA* and other downstream-located conjugation genes. Thus, TraA represses its own transcription together with other conjugation genes, but how it is induced or what is standby mode of this plasmid is not very clear. However, qPCR analysis showed that pIP501 *tra* genes are expressed during the entire growth cycle of *E. faecalis* (80).

pIJ101

Plasmid pIJ101 was originally isolated from mycelium forming *Streptomyces lividans ISP5434.* It is an 8,830-bp covalently closed circular (CCC) plasmid that encodes for a single gene, *tra*, which is essential for its intermycelial transfer (81). The Tra protein has homology with the DNA motor proteins SpoIIIE of *B. subtilis* and FtsK of *E. coli*, both of which were shown to translocate double-stranded DNA. Hence, conjugation in *Streptomyces* is a very distinctive process mechanistically and phenotypically from other bacteria (82, 83). Conjugative transfer takes place only on solid media during the early growth phase when *Streptomyces* grows as substrate mycelium. Thus, at other growth stages conjugation is switched off but the factor that induces conjugation at early growth phase is not known. The *tra* genes of many *Streptomyces* plasmids are under the transcriptional control of a GntR-type repressor (TraR/Kor). Since unregulated expression of *tra* is toxic, *tra* represents a kill function. In the case of plasmid pIJ101, KorA is the main repressor of conjugation and it represses transcription of *tra/kil*A and of *kor*A by binding to the overlapping divergent promoter regions (84) (see Fig. 5E). Consequently, KorA controls its concentration by autoregulating its own transcription. Secondly, repressor KorB binds to the *kil*B promoter 50-fold more efficiently than to the *kor*B promoter and inhibits the transcription from both the promoters (85). Upon receiving a copy of pIJ501, the plasmid can perform rapid intramycelial transfer of the plasmid to other cells. This intermycelial spread of the plasmid is called spreading function and is done by spreading genes, *spd* (indicated in yellow colour in Fig. 5E). Thus, Tra protein together with spreading genes are involved in pIJ101 plasmid transfer (81).

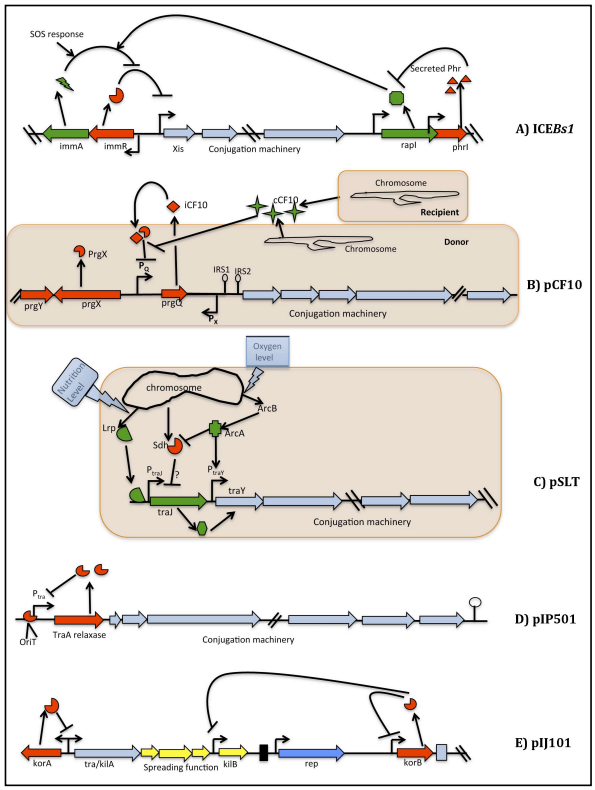

Figure 5. Schematic representations of regulatory circuits controlling conjugation genes present on prototype conjugative elements. Genes are represented by coloured arrows. Colour codes are used to indicate if the encoded gene product inhibits (red) or activates either directly or indirectly (green) the expression of the conjugation gnes (blue). **a)** ICE*Bs1* of *B.subtilis*. Expression of the conjugation genes are controlled at two levels: the SOS response and RapI. The system is maintained in a default "OFF" state by binding of a repressor,ImmR, to the promoter driving expression of the conjugation genes. Repression is relieved under conditions that induce the SOS response. Repression is also induced upon overexpression of RapI and the activity of RapI is regulated by Phr*I. RapI-mediated activation of conjugation genes requires the protease ImmA whose gene is located downstream of ImmR. **b)** *Enterococcus faecalis* plasmid pCF10. The ratio of inhibiting versus activating peptide (iCF10 [plasmid-encoded] to cCF10 [chromosome-encoded] determines the fate of conjugation. iCF10-PrgX complex repress the P_Q promoter. Recipient-produced cCF10 competes with iCF10 causing a conformational change resulting in derepression of the PQ promoter, which leads to increased synthesis of iCF10 together with conjugation genes. Accumulation of iCF10 returns the system to the default "OFF" state. **c)** *S. enterica* plasmid pSLT. The chromosomal encoded succinate dehydrogenase protein (Sdh protein, red symbol) represses traJ transcription by unknown manner, and hence mating. The host-encoded ArcA and Lrp proteins activate conjugation. **e)** *S. agalactiae* plasmidpIP501**.** traA, the first gene of conjugation operon encodes for DNA relaxase and regulates conjugation by repressing promoter P_{tra}. **e)** *S. lividans* plasmid pIJ101. Transcription of the crucial conjugation Tra protein is regulated by the KorA repressor. Another repressor, KorB, regulates the genes involved in intramycelium spreading of the plasmid (yellow arrows) via transcription of *kilB*.

Conclusion:

In the above section, we learned about Horizontal gene transfer and its role in bacterial evolution especially via conjugation. We know that good understanding of the process of conjugation and its transcriptional regulation can provide insights into bacterial evolution. These findings will also have important socioeconomic, medical and biotechnological implications. For example, it may provide valuable information to help control the explosive global spread of antibiotic resistance, and it may form the basis to construct tools to modify clinically or industrially important bacteria. We also noticed that a lot of study has been done for various plasmids present in Gram-negative bacteria but very less is known about plasmids from Gram-positive bacteria. To increase our understanding about conjugation system and its regulation in Gram-positive bacteria, we used a native plasmid from *Bacillus subtilis*. Below brief description about *B. subtilis* and the plasmid pLS20 is given.

1.4 *Bacillus subtilis*

The genus *Bacillus* includes both non-pathogenic and pathogenic species. Two pathogenic *Bacillus* species are considered medically significant: *B. anthracis,* which causes anthrax, and *B. cerecus,* which causes food poisoning. Other *Bacillus* species *B. thuringiensis* produces insecticidal endotoxins that are used to control insect pests and hence have an agricultural importance. Many species of *Bacillus* can secrete large amounts of important enzymes for example; *B. amyloliquefaciens* is the source of natural antibiotic protein barnase (a ribinuclease), α–amylase, proteinase subtilisin, and *Bam*HI restriction enzyme. Thus genus *Bacillus* consists of many species having industrial, medical and agricultural research importance

Bacillus subtilis in particular has been intensely studied over many years and, as a consequence, it is presently the best-characterized Gram-positive bacterium (86). The strong interest of the scientific community for *B. subtilis* was prompted by its developmental processes, like development of the competent state for DNA-binding and uptake, and formation of spores i.e. sporulation. The high amenability for genetic engineering and large-scale fermentation made *B. subtilis* an organism of choice for industrial applications (87). Whole genome sequencing of *B. subtilis* represented an enormous technology push (88, 89), which was followed up by genome-wide gene function analysis (90). Recently, *B. subtilis* was shown to be an excellent model

organism for Systems Biological analyses on gene regulation under different growth conditions (91, 92). Strikingly, whereas so much study has been done on *B. subtilis*, still little is known about their plasmids and even less about conjugation systems present on them. The main reason for this is that most of these studies were performed using the highly transformable *B. subtilis* strain 168 that does not harbor any endogenous plasmid. However, about 20% of natural isolates of *B. subtilis* are estimated to contain a plasmid (93). In addition, it has recently become clear that *B. subtilis* are gut commensals in animals and humans (94). This means that *B. subtilis* plasmids may play an important role in HGT at various levels and this warrants a better understanding of them.

1.5 Plasmid pLS20

pLS20 was first isolated from *B. subtilis* natto strain IF03335 and a physical map with several type II restriction enzymes was constructed by Tanaka *et al.* (95). Koehler and Thorne have shown that pLS20 is self-transmissible by conjugation and that it can also mediate interspecies transfer of mobilizable plasmids (96). pLS20 was the first reported example of a fertility plasmid in a *Bacillus* species other than *B. thuringiensis* (96). Itaya et al. labeled pLS20 with a chloramphenicol resistance marker, pLS20cat, and showed that transmission of pLS20cat occurred during exponential growth stage. Transfer efficiency reported was one order of magnitude higher (4.01×10^{-4}/donor/15min) in liquid media than transfer frequency reported for mating on solid media (1.3×10^{-5}/donor/15min) (14).

A 1.1kb region of the pLS20 contains all the information required for autonomous replication. This 1.1 kb origin region is located between two divergently transcribed genes, denoted *orfA* and *orfB,* both of which are not required for replication. The replication mechanism of pLS20 could not be classified to the Rolling Circle Mechanism (RCM) because of the lack of the characteristic features of RCM in the replicon of pLS20. Rather, the data obtained supports for that replication occurs via Theta-type Mechanism. The minimal replicon of pLS20 does not encode for a classical Rep protein and its replication is independent of PolI, both these characters are important for the classification of Theta type replication. Based on this and other data, pLS20 replication has been classified as a new class of theta replicons (97).

pLS20 contains an alp7A operon which has been shown to stabilize pLS20 plasmid (98, 99). Two genes *alp7*A and *alp7*R are present in the alp7A operon of the pLS20. Alp7A is a bacterial actin and its function requires that it assemble into filaments that treadmill and exhibit dynamic stability. Alp7R is a negative regulator and presumably

represses transcription of the entire operon, including *alp7R* gene. The negative feedback loop ensures that the appropriate amount of Alp7R is present in the cell to promote formation of normal dynamic filaments. Disruption of this feedback loop leads to aberrant filament formation, perturbation of plasmid segregation, and cell death (99). One study showing localization of conjugation machinery of pLS20 has been published (100). Recently, global affect of pLS20 on *B. subtilis* during stress has been studied (101).

2. Objectives

Objectives:

We chose pLS20cat plasmid to study regulatory circuit of conjugation. The following objectives were assigned to be addressed in this thesis.

1. Sequencing and annotation of pLS20cat plasmid.
2. Characterize the factor/s responsible for inhibition of competency in the presence of pLS20.
3. Determine the conjugation kinetics of pLS20
4. Characterize the circuit needed to regulate conjugation genes of pLS20cat

3. Materials and Methods

Bacterial strains, media, oligonucleotides and peptide

Escherichia coli and *B. subtilis* strains were grown in Luria-Bertani (LB) medium or on 1.5% LB agar plates (102). When appropriate, media or agar plates were supplemented with the following antibiotics: ampicillin (100 μg/ml), erythromycin (1 and 150 μg/ml for *B. subtilis* and *E. coli*, respectively), chloramphenicol (5 μg/ml), spectinomycin (100 μg/ml), kanamycin (10 μg/ml). Transformants were selected on LB agar plates with appropriate antibiotics. For sporulation experiments, *Bacillus* strains were grown in Schaeffer´s medium (103). Strains and plasmids used are listed in Supplementary Table 2 and Supplementary Table 3, respectively. *B. subtilis* strains are all isogenic with *B. subtilis* strain 168 (Bacillus Genetic Stock Centre code 1A700). Oligonucleotides used (Isogen Life Sciences, The Netherlands) are listed in Supplementary Table 4. Phr*$_{LS20}$ peptide was synthesized by the Proteomics department of our Institute.

Plasmid pLS20cat isolation, sequencing and sequence analysis

Plasmid pLS20cat used for sequencing was isolated by isopycnic cesium chloride ethidium bromide gradients (102). Two different approaches were used to clone pLS20cat fragments in pBluescript KS+ (pBSK+) vector. In the first approach, HindIII-digested fragments of pLS20cat were cloned in the corresponding site of pBSK+. In the second approach, purified pLS20cat DNA was partially digested with Sau3A and then fragments ranging in size between 0.6 and 2kb were isolated from agarose gels. Next, these fragments were clone in pBSK+ vector. To avoid cloning of multiple fragments in the same vector, the following strategy was followed. The pLS20cat Sau3A fragments were treated with Klenow enzyme in the presence of GTP and ATP nucleotides. Similarly, SalI linearized pBSK+ was treated with Klenow enzyme in the presence of nucleotides CTP and TTP. Next, a ligation mixture of these partially filled-in fragments was used to transform competent E. Coli XL1-Blue cells. Finally, plasmid DNA was isolated from 100 white transformants grown on IPTG/X-gal/ampicillin complemented agar plates and the borders of the cloned inserts were sequenced using universal primers M13Rev ad M13Fw-21. The DNA sequences obtained were then used to design primers, which were used in a subsequent round of DNA sequencing using maxi-prep purified pLS20cat as template DNA. pLS20cat was then sequenced to completeness in additional rounds by the primer walking strategy. Sequencing was performed by capillary electrophoresis on an ABI PRISM 3730x1 equipment (Applied Biosystems) with Big-Dye Terminators v3.1 at the Genomic Unit “Antonia Martin Gallardo” of the “Foundation Parque Cientifico Madrid (FPCM)”.

DNA sequences were analyzed and assembled using the Lasergene software package from DNASTAR, Inc (Madison, WI, USA). Putative genes were initially identified using the PRODIGAL program (104) and subsequent checked manually. Additional DNA analyses were performed using Sci Ed Central Clone manager Profession Suite (version 6) in combination with the online tools of Blast (http://blast.ncbi.nlm.nih.gov/Blast.cgi) and Expasy (http://www.expasy.org/).

Transformation

E. coli cells were transformed using standardized methods (102). For standard *B. subtilis* transformations, competent cells were prepared as described by Bron (1990). For making knockout version of pLS20cat, high competency protocol was used as described by Zhang and Zhang (105).

Construction of plasmids and strains

DNA techniques were performed using standard molecular methods (102). All enzymes used were purchased from New England Biolabs, USA. The correctness of all constructs was verified by sequence analysis.

Strain PKS1 containing rok_{LS20} driven by IPTG inducible P_{spank} promoter at the chromosomal *amyE* locus, was constructed as follows. First, the DNA region coding for rok_{LS20} and its 38 bp upstream region including a putative ribosomal binding site (RBS) was amplified from plasmid pLS20cat by polymerase chain reaction (PCR), using primers LSRok_FwHind and LS_Rok_RevXbaI. The purified PCR product was cleaved with *Hin*dIII and *Xba*I, and inserted into the *Hin*dIII and *Nhe*I sites of vector pDR110 (a gift from D. Rudner) to produce pPSrok. Next, the P_{spank}-rok_{LS20} cassette was placed at the *amyE* locus on the *B. subtilis* chromosome by transforming competent *B. subtilis* 168 cells with pPSrok and selecting transformants for spectinomycin resistance. Double crossover events were confirmed by the loss of the *amyE* functionality. The same protocol was used to clone rco_{LS20} and rap_{LS20} under IPTG inducible P_{spank} promoter and placed at amyE locus on the *B. subtilis* chromosome.

In plasmid pPKS26 *rapI* is placed under the control of the $P_{hyspank}$ promoter. This plasmid was constructed by first amplifying a *rapI* containing DNA fragment by PCR using oligos oGR85 and oGR86 and *B. subtilis* 168 DNA as template. Next the PCR fragment was digested with *Nhe*I and *Sph*I and cloned in vector pDR111 digested with the same enzymes. Finally, the $P_{hyspank}$-*rapI* construct was placed at *amyE* locus of the ICE*Bs1* negative strain PY79 by using plasmid pPKS26 to transform competent PY79

cells resulting in strain PKS139.

To construct a translational fusion of rok_{LS20} with *yfp* encoding yellow fluorescent protein, the rok_{LS20} gene lacking its stop codon was amplified from plasmid pLS20cat using primers oPKS3 and oPK11. Next, the purified PCR product was digested with *Eco*RI and *Kpn*I and the fragment was inserted into the *Eco*RI-*Kpn*I sites of pSG4924 to produce clone pPKSN7. Strain PKS44 was obtained by transforming competent cells of strain PKS21 with pPKSN7.

The following strategy was followed to construct a translational fusion of rok_{LS20} with *his(6)*. The rok_{LS20} gene was amplified from pLS20cat by PCR using primers oGR8 and oGR9. The purified PCR product was digested with *Nde*I and *Xho*I and cloned into the vector pET28b(+) digested with the same restriction enzymes to produce plasmid pHis-Rok_{LS20}.

The following strategy was followed for constructing *B. subtilis* strain containing a copy of *lacZ* fused to the rco_{LS20}-gene 28 intergenic region. Almost 570 bp upstream sequences of gene 28 of pLS20cat plasmid were amplified using primers Prom28UpBam and Prom28UpHind. The purified PCR product was cleaved with *Hind*III and *Bam*HI, and inserted into same sites of the vector pDG1663 to produce pDGP28. Next P_c-*lacZ* construct was placed at *thr*C locus at the *B. subtilis* chromosome by transforming competent *B. subtilis* 168 cells with pDGP28 resulted in erythromycin resistance strain PKS3. Double crossover events were confirmed by the loss of spectinomycin resistance.

For constructing the rok_{LS20} knockout plasmid pLS20rok, we amplified the 2 kb upstream and downstream regions of rok_{LS20} using primer-sets oPKS16/oPKS17Pst and oPKS18Eco/oPKS19 respectively. The PCR fragments were cleaved with *Pst*I and *Eco*RI respectively, and used in a ligation mixture together with a *Pst*I-*Eco*RI DNA fragment encompassing the neomycin resistance marker obtained from plasmid pBEST501. This ligation mixture was then used to transform 'supercompetent' PKS56 cells that harbour pLS20cat and contain a P*xyl*-*comK* construct. Replacement of rok_{LS20} on pLS20cat by neomycin resistance gene was confirmed by PCR. The same protocol was used to construct derivatives of pLS20cat in which the rap_{LS20}, phr_{LS20} *or* rco_{LS20} genes were replaced by an antibiotic resistance marker. Transcriptional fusions P_{comK}-*lacZ* and P_{comG}-*gfp1*, and the Δrok_{Bs168} construct were introduced in the *B. subtilis* 168 background by transforming competent cells of strain 168 with total DNA of strains 8G33, 8G5-ComG-gfp and WKD1039.

Strains containing pLS20cat were obtained by conjugating pLS20cat using PKS11 as donor cells.

β-Galactosidase activity assays

Overnight cultures were diluted 100-fold into fresh medium and samples were taken at 45 min intervals for optical density reading (OD_{600}) and determining β-galactosidase activity as described previously (Miller, 1982).

Fluorescence microscopy

Cells were grown in competence media, or in LB with 0.02% xylose were placed on agarose pads (1.2%) (Harwood and Cutting, 1990). DAPI (4′,6-diamidino-2-phenylindole) (Sigma) and membrane dye FM95 (Invitrogen) was added to cells 5 min prior to visualization. Images were acquired using a Nikon Eclipse Ti-U inverted epifluorescence microscope and a QImaging Rolera EM-C2 EM-CCD Camera under 100¥ phase oil objective, and were processed using MetaMorph software. GFP expressing cells and total number of cells were counted by ImageJ cell counter program. TIFF images were further processed in Adobe Photoshop CS2.

Conjugation assays

Unless specified otherwise, conjugation was carried out in liquid medium as described by Itaya et al. (14). Thus, for standard conjugation experiments, overnight cultures of donor and recipient cells, grown in the presence of appropriate antibiotics, were diluted 25 fold in fresh 37 ^{0}C pre-warmed LB medium without antibiotics and grown for 2.5 h in shaking (125 rpm) water bath. Next, 200 μl of both donor and recipient cells were mixed in 2.5 ml eppendorf tube and incubated for 15 min at 37 ^{0}C without shaking to permit conjugation. Finally, appropriate dilutions were plated on LB agar plates supplemented with proper antibiotics to select either for transconjugants or for donor cells. When conjugation efficiencies were determined as a function of growth, overnight cultures were diluted to an OD_{600} of 0.01. Next, donor and recipient cells were grown separately (180 rpm) and 200 μl of the donor and recipient cultures were withdrawn at different times and proceeded as described above. Growth was followed by measuring OD_{600} at regular intervals. In order to study the effect on conjugation of over-expression of a given gene placed under the control of the inducible P_{spank} promoter, IPTG was added to prewarmed LB medium used for inoculation of the overnight grown cultures. Unless mentioned otherwise, IPTG was added to a final concentration of 1 mM.

His(6)-Rok$_{LS20}$ purification

An overnight culture of *E. coli* BL21 (DE3) carrying pHis-Rok$_{LS20}$ was used to inoculate (100-fold dilution) 1 l of fresh LB medium containing 30 mg ml^{-1} kanamycin and incubated at 37°C with shaking. At OD_{600} of 0.5, His(6)-Rok$_{LS20}$ was induced with 1 mM IPTG at 37°C and growth was continued for 1 h. Cells were collected by centrifugation and washed in 1/10 vol. of buffer A (250 mM NaCl, 10 mM MgCl2, 20 mM Tris-HCl pH 8, 7% glycerol, 10 mM imidazole, 1 mM b-mercaptoethanol). After cells were centrifuged and re-suspended in 1/3 vol. of buffer A they were lysed by sonication followed by DNase I treatment for 30 min at 4°C. Next, the lysate was centrifuged twice (15 k, 30 min) and the supernatant was collected, mixed with 1 ml of nickel NTA agarose beads equilibrated with buffer A. The mixture was incubated end-over-end for 1 h at 4°C then packed into a column. The column was washed with extensive amounts (> 50 column volumes) of buffer A containing increasing concentrations (10, 20, 30, 50 and 100 mM) of imidazole. Next, the His(6)-Rok$_{LS20}$ protein was eluted in eight fractions of 1 ml of buffer A containing increasing concentration of imidazole (stepwise from 250 to 500mM). All fractions were analysed by SDS-PAGE and only the fractions with > 95% purity were pooled, dialysed against buffer B (20 mM Tris-HCl pH 8.0, 1 mM EDTA, 250 mM NaCl, 10 mM MgCl2, 7 mM b- mercaptoethanol, 20% v/v glycerol) and stored in aliquots at -80°C. Protein concentrations were determined by Bradford assay.

Gel retardation

The *comK* and *comG* promoter regions were amplified by PCR using primer sets PcomK1/PcomK2 and PcomG1/ PcomG2 respectively, with 168 chromosomal DNA as tem- plate. The resulting PCR fragments were purified and equal concentrations (300 nM) were incubated on ice in binding buffer [20 mM Tris HCl pH 8, 1 mM EDTA, 5 mM $MgCl_2$, 0.5 mM DTT, 100 mM KCl, 10% (v/v) glycerol, 0.05 mg/ml BSA] without and with increasing amounts of purified His(6)- Rok$_{LS20}$ in a total volume of 16 ml. After careful mixing, samples were incubated for 20 min at 30°C, placed back on ice for 10 min, then loaded onto 2% agarose gel in 0.5 TBE. Electrophoresis was carried out in 0.5 TBE at 50 V at 4°C. Finally, the gel was stained with Ethidium Bromide (EtBr), destained in 0.5 TBE and photographed with UV illumination.

RNA isolation and RNA sequencing

Total RNA was isolated from late exponentially growing cells by using RNeasy Mini Kit from Qiagen according to manufacturer´s protocol. RNA protect solution provided by Qiagen was used to ensure the integrity of RNA during isolation and also to stop transcription at given time points. RNA was treated with DNAseTurbo (Ambion) to remove possible traces of contaminant DNA. Between 5 to 15 µg of total RNA was subjected to rRNA removal using RiboZero (Epicentre, either Gram-positive specific or metabacteria-specific) following the manufacturer instructions to obtain 150-250 ng of rRNA-depleted RNA. Next, RNA of each sample was used to prepare cDNA libraries using a procedure that preserves information about transcript's direction (ScriptSeq mRNA library prepration kit, Illumina compatible; Epicentre)(106). As specified by the supplier, samples were fragmented for 5 min at 85 ^{0}C and subsequently barcoded so that they could be run in combination.

After library prep, samples were titrated by quantitative PCR, pooled and bound at a final concentration of about 10 pM to an Illumina SR-flowcell using a Cluster Station apparatus (Illumina). Libraries were then run on a GAIix equipment (Unidad de Genómica, Parque Científico de Madrid) by SBS under a single-read 1x75 protocol. Quality filtering was performed automatically according to Illumina specifications and fastq files generated.

Bioinformatic analysis of RNAseq data

Data set. The analyzed data set was constituted by five *B. subtilis subsp. subtilis* str. 168 and plasmid pLS20cat samples corresponding to four different experimental conditions, with a total of 56,439,165 single end reads of 36 nt length in FASTQ format. Data were analyzed using the standard bioinformatic analysis workflow of a RNA-seq experiment.

Reads quality. A preliminary analysis of the quality of the reads was performed using FastQC, a Java tool with graphic interface (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Percentages between 93.02 % and 93.24 % of all bases had a quality score of 30 or higher (probability of incorrect base call of 10^{-3} or lower) and between 85.43% and 85.85% of all bases had a quality score of 35 or higher (probability of incorrect base call of $3{\cdot}10^{-4}$ or lower), being 40 the maximum score quality reported in FASTQ format (probability of incorrect base call of 10^{-4} or lower). Because of the high quality it was not necessary to process the reads by filtering or trimming them.

Alignment: The reads were mapped to the published *B. subtilis subsp. subtilis* str. 168 and plasmid pLS20cat reference genomes using Bowtie software(107) with the following parameters. Maximum allowed number of mismatches 3, input qualities are Phred+33 (108), and the "--best" option was switched on, ensuring that reported

alignments are "best" in terms of chosen criteria (allowed number of mismatches), and that alignments are reported in best-to-worst order. Of the total reads, a percentage between 92.48 and 98.51 % could be mapped to the reference genome with 79 to 106-fold sequencing coverage across the entire genome. Unmapped reads were searched in UniVec database using BLAST(109). UniVec is a database that contains DNA sequences of cloning/expression vectors, adapters, linkers, and primers that are commonly used in the process of cloning and sequencing nucleic acids (http://www.ncbi.nlm.nih.gov/VecScreen/UniVec.html). This database was used to identify such contaminating sequences from the unmapped reads. Of total reads, percentages between 0.18 % and 0.56 % were assigned to UniVec database sequences, revealing very low levels of vector contamination. Unmapped reads were discarded for further analysis. Out of the total of 56,439,165 reads, 1,596,385 (2.83 %) mapped to the pLS20cat genome, which were used to calculate expression levels of individual pLS20cat genes under the different conditions.

Expression levels: The alignment files were processed using EpiCenter software (http://www.niehs.nih.gov/research/resources/software/biostatistics/epicenter/), an analysis tool of genome-wide mRNA-seq or ChIP-seq data for detecting differentially expressed genes(110) .

Plasmid pLS20cat expression levels were additionally used to draw a heat map that graphically shows the expression levels of "wild type" experimental conditions (left lane on Figure R12). In addition, the heatmap shows the differences in expression of pLS20cat genes when Rco_{LS20} or Rap_{LS20} were ectopically expressed (middle and right lanes Figure R12, respectively) compared to the wild type situation by using Matrix2png software (http://www.chibi.ubc.ca/matrix2png/) (111).

Computer-assisted analysis

Protein blast (blastP and psi-blast) searches (http://blast.ncbi.nlm.nih.gov/Blast.cgi) were performed for each ORF of pLS20cat to gain insights in the function of the protein encoded by these ORFs. Alignments of the primary amino acid sequences of Rok and Rco homologues were made using the ClustalW2 program (http://www. ebi.ac.uk/clustalw) and confirmed using the CloneManager 9 professional suite software (Sci-Ed Software, Scientific and Educational Software Cary NC, 27513 USA). Adobe Photoshop CS2 and Adobe Illustrator were used for creating figures and arts work. The Excel program was used to create graphics.

4- Results

4.1 In-silico analysis of pLS20cat sequences

4.1.1 Sequencing, annotation and general features of pLS20cat

For several reasons as outlined above, it is important to improve the knowledge of conjugative elements of Gram-positive bacteria. Since pLS20 is the only conjugative plasmid of *B. subtilis* described so far, we chose this plasmid for our studies. At the start of the project, only two small regions of pLS20cat were sequenced (97). Therefore, the first objective was to determine the entire sequence of pLS20 and annotate its sequence.

The entire pLS20cat plasmid was sequenced by the dideoxychain termination method using a combination of sequencing PCR fragments and primer walking strategy (see M&M). Plasmid pLS20cat has a total size of 65,774 bp and its (G+C) content is 37.8 %. pLS20cat contains 92 open reading frames (ORFs) with a size of at least 30 codons and that are preceded by a putative Ribosomal Binding Site (RBS). The ORFs were detected using programs like CloneManager and Prodigal (Prokaryotic Gene Prediction Program). The deduced protein sequences of all 92 ORFs were tasted for homology with protein sequences present in variable databases using Blast. All this information was used to generate a circular map of plasmid pLS20cat using the program SnapGene viewer (http://www.snapgene.com/), which is presented in Fig. R1. A summary of the characteristics of identified genes/ORFs including their putative RBSs of plasmid pLS20cat is given in Supplemental Table S1.

The identified features combined with information of the related plasmid p576 (112) were used to categorize regions and ORFs/genes of pLS20cat into five modules, (i) plasmid partitioning functions, (ii) origin of replication, (iii) *rap-phr* cassette, (iv) putative conjugative transfer operon with flanking gene encoding putative repressor, and (v) genes/ORFs of unknown functions. A schematic map of pLS20cat is presented in Fig R1 in which the five modules are indicated with different colours.

(i) Partitioning functions:

Plasmids are extrachromosomal DNA molecules that generally do not carry essential genes and hence may be lost over time. In order to be maintained stably over time, plasmids must either (a) provide selective advantage to the cell, (b) possess effective systems that ensure faithful segregation and/or systems that otherwise prevent plasmid loss.

Presently, we do not have an indication that pLS20 provides a selective advantage to its host. However, pLS20 appears to carry at least two systems that probably contribute to its segregational stability. These two systems are (a) true partitioning functions, and (b) post-segregational killing functions (PSK).

During the annotation of pLS20cat, we identified one true partition system, name Alp7A-Alp7R (ORFs 24c and 23c, respectively), which has been investigated by others (99). In addition, we identified a gene 62 that may form part of a post-segregational killing function.

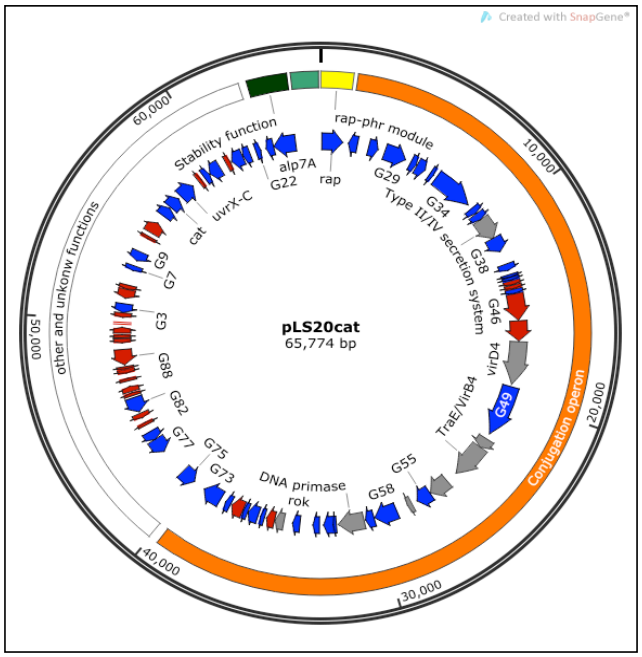

Figure R1: **Genetic map of pLS20cat.** Outer, middle, and inner circle indicate bps, different modules, and gene position, respectively. Gene 1 (shown in the module of other and unknown functions) corresponds to the homologue of gene 1 of the related *Bacillus pumilus* NRS576 plasmid p576(112). The position, length and orientation of the (putative) genes are indicated with arrows. The colours of the arrows indicate their function based on homology of the deduced gene product with proteins present in databases (see Table S1). Grey arrows, conjugation genes; blue arrows, genes for which a function could be attributed based on homology that is different than conjugation; red arrows, putative genes of unknown function. In the middle ring, the origin region, conjugation operon, *rap-phr* cassette and other (unknown) functions are indicated with green, orange, yellow, and white boxes, respectively. Number for each gene is preferred rather than putative names (except few) to avoid complications in the map.

(ii) Replication region:

A 1.1 kb region of pLS20cat, sufficient to sustain autonomous replication, was identified before and its characterization led to define it as a new class of theta replicons based on its unique features (97). The origin region does not encode for a gene encoding a replication protein typically present on most theta-replicating plasmids. The single ORF encoding a putative protein of more than 80 amino acids,

named Marta (82 codons) does not contain a clear potential RBS, and its deduced protein sequence does not show significant homology with proteins present in available databases, suggesting that it does not constitute a gene. The region contains a few (im)perfect DnaA boxes and stem loops. A possible role of these regions in replication is not clear at the moment.

(iii) The *rap-phr* module:

The deduced protein sequence of pLS20cat ORF25 (368 codons) shares significant homology with members of the Rap family of proteins. ORF25 is followed by a small ORF (44 codons) that probably encodes a precursor of a secreted Phr molecule. Based on this, we designated ORF25 and ORF26 as rap_{LS20} and phr_{LS20}, respectively. A functional analysis of the rap_{LS20}-phr_{LS20} module will be described later in this thesis.

(iv) Putative conjugative transfer operon

Based on the following features, the region encompassing putative genes 28 to 74 is likely to constitute a large conjugation operon. First, ORF28 up to ORF74 are all transcribed in the same direction. Second, some pLS20cat genes show significant homology with essential conjugation genes present on the other conjugative elements (see Table S1 for an overview). And third, essential conjugation genes are not found outside of this region. A succinct overview of the conjugation genes identified on pLS20cat is given below.

Processing of the DNA substrate for conjugation is initiated by formation of a relaxasome, i.e. the binding of a relaxase and one or more accessory factors to the origin-of-transfer (oriT). The relaxase then produces a site-specific nick in the DNA strand that is destined to be transferred. After nicking, the relaxase remains bound to the 5' end of the nicked DNA, and "delivers" the substrate to the translocation channel by direct interaction with a conserved ATPase named Type IV coupling protein (T4CP, aka TraG/TrwB/VirD4 family of coupling proteins). In the recipient cell, the relaxase catalyzes the recircularization of the transferred single-stranded DNA molecule and may also participate in double-strand DNA synthesis (54, 113). ORF59 of pLS20cat encodes for a putative relaxase (see Table S1). The localization of *ori*T of pLS20cat is being investigated by others in the lab.

The deduced protein sequences of ORF48 shows similarity with VirD4 component (see Table S1) and most likely encodes the T4CP of pLS20cat. The T4CP physically interacts with the translocation channel through which the DNA and some proteins are

transported from donor to the recipient cells. Thus, T4CP forms an interface between the relaxosome and the translocation channel.

An important component of conjugation machinery is the translocation channel, which is comprised of the mating-pair formation (mpf) proteins. Putative proteins encoded by ORF37, ORF52 and ORF54 show homology to different functional classes of mpf proteins (see Table S1). There are several ORFs in the conjugation operon of pLS20 that are predicted to encode proteins containing one or more trans-membrane spanning domain (see Table S1). Probably these proteins contribute in various ways to the channel formation and activity.

During conjugation, DNA is transferred to the recipient cell in its single-stranded form (ssDNA), which is vulnerable to degradation, can be protected by Ssb (single-stranded DNA binding) proteins (113). The deduced protein sequence of pLS20cat ORF66 has 67% sequence identity with *B. subtilis* chromosomal Ssb.

The deduced protein sequences of several ORFs in the putative conjugation operon do not show significant homology to classical conjugation genes. However, a few of them do show homology to other genes present in databases. For example, the deduced protein sequence of ORF64 shows homology with Rok, which is encoded by the *B. subtilis* chromosome and which we renamed Rok_{Bs168}. Rok_{Bs168} inhibits the development of natural competence (114). The presence of rok_{Bs168} homologue on pLS20cat was a surprise. A detailed characterization of pLS20cat ORF64 is presented in this thesis (Part 2).

ORF27 is divergently orientated to the flanking ORFs of the putative conjugation operon and shows homology to Xre-type transcriptional regulator proteins. Thorough analysis of ORF27c has revealed that it forms an important part of regulatory circuitry of pLS20cat conjugation. Detailed discussion about regulation and role of ORF27 is given later on in this thesis.

(v) Other and unknown function

A few ORFs show homology to other genes present in databases. For example, ORF82 shows significant homology to ArdC-type of antirestriction proteins. Antirestriction proteins are inhibitors of type I restriction-modification enzymes and are found on various conjugative plasmids, ICEs and bacteriophages {for review see, (115)}. For some antirestriction proteins evidence exists that during conjugation they are transferred along with the ssDNA molecule into the recipient cell to protect the plasmid from digestion by type I restriction enzyme. It was reported that plasmids carrying these genes could enter into a host more efficiently (116). Based on these data, it is plausible that ORF82 help for the establishment of pLS20cat after entering the recipient.

ORF14 encodes for Chloramphenicol acyltransferase (cat), which provides resistance against chloramphenicol antibiotic. This gene has been cloned by Itaya et al., 2006 (14) at an unique SalI site present on pLS20.

ORFs 13 and 15, both show significant homology with N-terminal and C-terminal of Uvr repair protein, respectively. Thus ORF13 and ORF15 were part of one single gene of pLS20. Presence of gene for Uvr repair on pLS20 might be advantageous for its host adaptation during exposure of ultraviolet rays.

Many ORFs of pLS20cat could not be assigned any function, as they did not show any homology to present database.

4.2 Presence of plasmid pLS20cat inhibits development of natural competency of *Bacillus subtilis* 168

Under certain growth conditions *B. subtilis* develops natural competence. In this state, *B. subtilis* cells can take up exogenous DNA and incorporate it in its genome by homologous recombination. Thus, the development of natural competence facilitates easy genetic modification of the genome and its native plasmids. In order to modify plasmid pLS20cat, we prepared competent cells of wild type 168 a strain harbouring pLS20cat (PKS11) and transformed it with desired DNA. However, no or very few transformants were obtained while control experiments using the isogenic strains lacking pLS20cat gave normal transformation efficiency. This result suggested that the presence of pLS20 interfered with the competence development. Finding out the underlying mechanism responsible for pLS20-mediated inhibition of competence is intrinsically interesting because it may shed light on the interaction between pLS20 and its host. In addition, it was important to overcome this competence inhibition in order to be able to genetically manipulate pLS20cat. This prompted us to study in some detail the mechanism underlying pLS20-mediated inhibition of competence.

4.2.1 Presence of pLS20cat inhibits competence of its host

The presence of pLS20cat inhibits competence development of its host, we performed additional transformation experiments using strains containing or lacking pLS20cat. The experimental procedure followed is given in M&M sections. As shown in Table R1 the efficiency of transformation obtained with strain PKS11 harbouring pLS20cat was almost 100 fold lower than the isogenic strain 168 (1A700) lacking pLS20cat.

Table R1. pLS20cat inhibits competence of its host.

Strain	Genotype	Transformation efficiency
1A700	Wild type (168)	$2.14x10^{-2}$
PKS11	168, pLS20cat	$2.01x10^{-5}$
PKS69	168,pLS20rok	$6.4x10^{-4}$

4.2.2 pLS20cat inhibits competence of *B. subtilis* 168 via the *comK* pathway

ComK is the master regulator for competence (117). Competence develops in cell in which sufficiently high level of ComK have accumulated to activate the expression of the down-stream competence genes (118). A reliable method has been described to monitor ComK production in single cells that is based on a transcriptional fusion between the ComK-dependent promoter of the *comG* operon (P_{comG}) and the gene encoding for green fluorescent protein (GFP) (119). High levels of ComK are required to activate the P_{comG} promoter. We used this system to study the effect of pLS20cat on ComK-mediated activation of the P_{comG} promoter in individual cells. Thus, we introduced pLS20cat into strain PKS16 that contains the P_{comG}-*gfp* construct resulting in strain PKS17. We then grew strains PKS16 and PKS17 in competent medium and determined the fraction of cells that developed competence by counting the number of

green cells under fluorescence microscope. The results obtained revealed that only 1.7% of the cells harbouring pLS20cat (PKS17) were fluorescent, compared with 14.5% of the cells lacking the plasmid (PKS16). This result together with the above ones confirms that the presence of pLS20cat affects the development of competence by lowering the fraction of cells expressing high level of ComK.

4.2.3 pLS20cat contains a putative gene homologous to *B. subtilis 168 rok*

The deduced protein sequence of ORF64 shows significant homology with the Rok_{Bs168} encoded by *B. subtilis* (see Fig. R2). Rok_{Bs168} inhibits natural competence by repressing ComK expression (114). We speculated that ORF64 of pLS20cat might be responsible for competence inhibition.

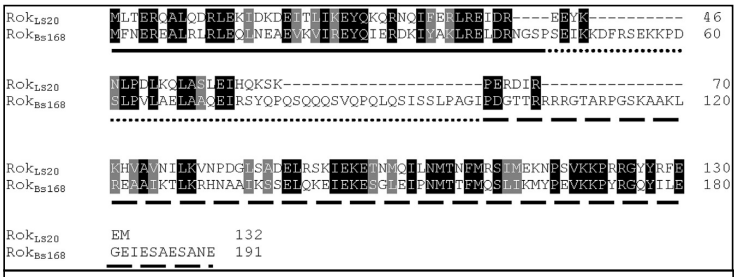

Figure R2 Alignment of the deduced protein sequences of putative Rok_{LS20} encoded by pLS20 and Rok_{Bs168} encoded by the chromosome of *B. subtilis* 168. Identical and conserved residues are shown against a black and grey background respectively. The N-terminal, middle, and C-terminal regions of Rok_{Bs168} as defined by Smits and Grossman (2010) are indicated with a solid, dotted and dashed lines, respectively.

Based on several *in silico* and functional analysis, Rok_{Bs168} has been divided into three distinct regions: (i) a conserved N-terminal region (residues (1-45), (ii) a less conserved central region (residues 46-95), and (iii) a conserved C-terminal DNA binding region (residues 96-191) (120). Most of the central region of Rok_{Bs168} is absent from the putative Rok_{LS20} (see Fig. R2). Consequently, Rok_{LS20} is smaller than Rok_{Bs168} (132 aa versus 191 aa). The C-terminal region of Rok_{Bs168} is required for DNA binding whereas the N-terminal and central regions enhance the binding activity and specificity for A+T-rich DNA (120). Rok_{LS20} lacks the central region and there are substantial differences between the C-terminal regions of Rok_{LS20} and Rok_{Bs168} (Fig. R2). This may be an indication that the two proteins have different DNA binding specificities and/or activities.

To study if the inhibitory effect on competence by pLS20cat was due to the activity of Rok_{LS20}, we constructed a derivative of pLS20cat, pLS20rok, in which rok_{LS20} was replaced by a neomycin resistance gene (see M&M). We compared the transformation efficiencies of *B. subtilis* strains lacking or harbouring either pLS20cat or pLS20rok. The transformation efficiencies of strain PKS11 (harbouring pLS20cat) were lower than

those obtained with strain PKS69 (harbouring pLS20rok). This result indeed indicates that the presence of an intact copy of *rok*$_{LS20}$ on pLS20cat affects competence development. However, the transformation efficiencies obtained with the plasmid free strain 1A700 were higher than those obtained for PKS69. This indicated that pLS20cat affects transformation efficiency also in a *rok*$_{LS20}$-independent manner. At present we do not know how pLS20 inhibits competence via *rok*$_{LS20}$-independent way, but it appears that *rok*$_{LS20}$ has a clear effect and we focused the way *rok*$_{LS20}$ inhibits competence.

4.2.4 Ectopic expression of Rok$_{LS20}$ inhibits competence of *B. subtilis* 168 by repressing *comK* expression

To study the effect of Rok$_{LS20}$ on competence, we constructed a strain, PKS1, which contains a single copy of a cassette in which *rok*$_{LS20}$ is under the control of the IPTG inducible P$_{spank}$ promoter and thus allowing conditional expression of Rok$_{LS20}$. Competence levels were determined for strain PKS1 and an isogenic strain lacking *rok*$_{LS20}$ (PS110) in the presence and absence of the inducer IPTG. The results obtained show that induction of *rok*$_{LS20}$ led to a large decrease in competence levels (>100-fold), indicating that Rok$_{LS20}$ inhibits competence (see Table R2).

Table R2. Ectopic expression of Rok$_{LS20}$ inhibits transformation efficiency

Strain	Genotype	IPTG	Relative transformation efficiency
PS110	168, amyE::P$_{spank}$- , spec	-	75
		+	95
PKS1	168, amyE::P$_{spank}$-*rok*$_{LS20}$, spec	-	100
		+	0.7
PKS22	168,amyE::P*spank-rok*$_{LS20}$,spec *rok*$_{Bs168}$::Cm, P$_{comG}$-*gfp*, Kan	-	254
		+	5.4
Experiments were done in triplicates and results are mean values of all three experiments			

The presence of pLS20cat decreases the number of cells expressing ComG-gfp (see above). We used the same strategy to determine if expression of Rok$_{LS20}$ causes the same effect. For this, we introduced P$_{comG}$-*gfp* fusion into strain PKS1 and analyzed cells of the resulting strain, PKS18, by fluorescent microscopy (see M&M). Representative images of PKS18 cells grown in competence medium with or without 1 mM IPTG are shown in Figure R3. Only about 1% of the cells expressed GFP when Rok$_{LS20}$ was overexpressed (Fig. R3B), while about 15% of the cells expressed GPF without induction of *rok*$_{LS20}$. These results indicate that Rok$_{LS20}$ acts negatively on the expression of ComK.

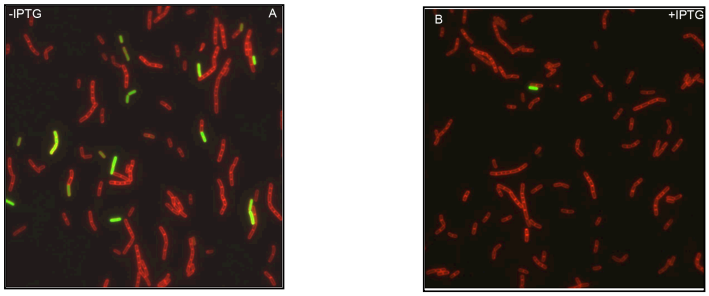

Figure R3. Rok_{LS20} limits the fraction of cells that develop competence by inhibiting *comK* expression. **A** and **B.** Fluorescence microscopy of PKS18 cells which carry a P_{comG}-*gfp* fusion construct together with an IPTG-inducible rok_{LS20} gene at the *amy*E locus grown without (**A**) or with 1 mM IPTG (**B**). Cells were stained with FM5-95 membrane dye to visualize the membranes. Percentages of fluorescent cells obtained in two independent experiments were 14.5% and 18.0% in the absence of IPTG, and 0.4% and 1.1% in the presence of IPTG respectively. At least 1000 cells were analysed for each culture.

To confirm this conclusion we studied the activity of the *comK* promoter (P_{comK}) directly using a P_{comK}-*lacZ* fusion. A cassette containing a transcriptional P_{comK}-*lacZ* fusion was introduced into strain PKS1 to give strain PKS66. β-galactosidase activities were determined for samples taken at different time points during growth from PKS66 cultures grown in competence medium (+/- IPTG). As expected, P_{comK} expression becomes activated near the end of the exponential growth phase when cells were grown in the absence of IPTG. However, very low β-galactosidase activities were obtained throughout growth when cells were grown in the presence of IPTG (Fig. R4). There results show that Rok_{LS20} inhibits *comK* promoter activity. Together, the results provide evidence that Rok_{LS20} inhibits expression of ComK, which affects the subpopulation developing competence.

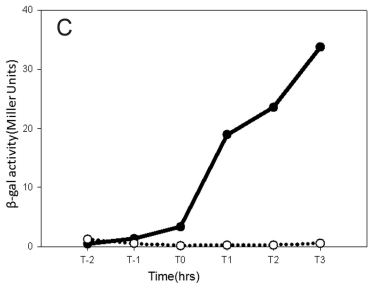

Figure R4. The *comK* promoter is repressed by Rok_{LS20}. Cultures of PKS66 cells were grown in competence medium in the absence (-) or presence (+) of 1 mM IPTG and β-galactosidase activities were determined of samples withdrawn at different times. Time is given in hours before or after the transition to stationary phase T0.

4.2.5 Rok_{LS20} acts at the *comK* transcription level and binds preferentially to the *comK* promoter

The cellular level of ComK is not only regulated at the transcriptional level, but also at the post-transcriptional level through modulation of the ClpCP protease system. ClpCP protease degrades ComK in the absence of ComS (117). To determine if Rok_{LS20} acts at the transcriptional or post-transcriptional level of *comK*, we constructed strain PKS82, in which the native *comK* locus has been inactivated but contains a xylose-inducible copy of *comK* at an ectopic position.

Consequently, ComK expression levels in this strain depend only on the activity of the P_{xyl} promoter, which is insensitive to Rok_{Bs168}. Therefore, competence of this strain would be affected only if Rok_{LS20} acted on ComK at the posttranscriptional level. Result obtained showed that overexpression of Rok_{LS20} in strain PKS82 had no effect on competence level ($1.63x10^{-3}$ with overexpression versus $2.98x10^{-3}$ without overexpression of rok_{LS20}). The insensitivity of the xylose-induced ComK production to the Rok_{LS20} provides compelling evidence that Rok_{LS20} does not act posttranscriptionally on ComK. Next, we confirmed these results using different approach. We tested whether Rok_{LS20} could act indirectly on *comK* expression via ComS, the small protein that prevents ComK degradation by the ClpCP protease. As the *comS* gene is part of the *srfA* operon (121, 122), we used a transcriptional *srfA-lacZ* fusion to examine the effect of the rok_{LS20} expression on the transcription of *comS*. As shown in Figure R5, the activity of P_{srf} was not significantly altered upon induction of rok_{LS20}, providing additional evidence that Rok_{LS20} does not act posttranscriptionally on ComK expression.

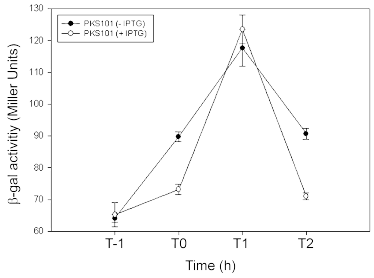

Figure R5. Overexpression of Rok_{LS20} does not affect P_{srf} expression. Cultures of strain PKS101 ($P_{spank\text{-}rok_{LS20}}$, $P_{srf\text{-}lacZ}$) were grown in competence medium complemented with or without IPTG at 37°C and ß-galactosidase activities were determined of samples withdrawn at different times. Time is given in hours before or after the transition to stationary phase T0.

The results presented above demonstrate that Rok_{LS20} inhibits P_{comK} activity and that it does not affect the stability of ComK by altering ComS expression. A likely scenario to explain these results is that Rok_{LS20} represses directly the *comK* promoter.

To analyze this scenario, we used gel-retardation experiments. For this, we constructed a his-tagged version of rok_{LS20} and purified the His(6)-Rok_{LS20} fusion protein by affinity chromatography. The results presented in the Figure R6 confirm that His(6)-Rok_{LS20} preferentially binds to the *comK* promoter. The result of the *in vitro* experiment

and the negative result of P_{srf}-*lacZ* experiment provide evidence that Rok_{LS20} acts at the promoter of *comK*.

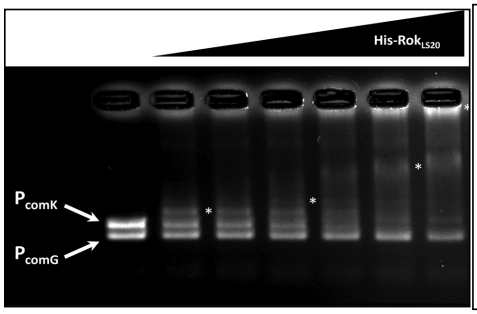

Figure R6. Rok_{LS20} binds preferentially to P_{comK}. Equal amounts of *comK* (303 bp) and *comG* (249 bp) promoter-containing fragments (300 nM) were incubated without or with increasing amounts of purified His(6)-Rok_{LS20} and subsequently run on a 2% agarose gel. Positions of *comG* and *comK* promoter fragments are indicated with arrows. Shifted bands are indicated with asterisks. His(6)-Rok_{LS20} concentrations were in the range of 7.2 to 2.8 mM.

4.2.6 Localization of the Rok_{LS20} inside host cell

To study the subcellular distribution of Rok_{LS20}, we constructed a strain PKS44, containing a xylose-inducible *rok*$_{LS20}$-*iyfp* fusion located at the *amyE* locus in a *rok*$_{Bs168}$ deletion background, and used fluorescence microscopy to visualize the location of IYFP-labeled Rok_{LS20} in living cells. Functionality of the translational fusion (Rok_{LS20}-IYFP) was confirmed by checking transformation efficiency in presence and absence of inducer (not shown).

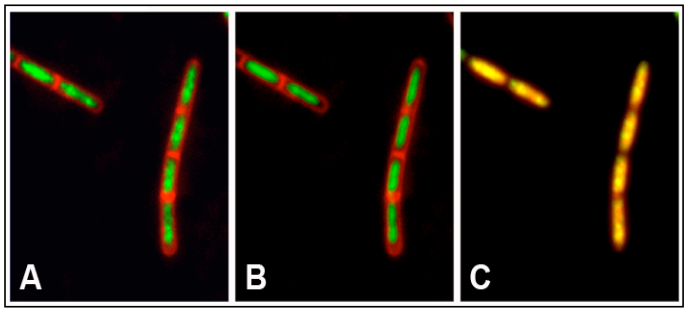

Figure R7. Rok_{LS20}-IYFP colocalizes with the nucleoid. Strain PKS44 (*rok*$_{LS20}$-*iyfp*) was grown to mid-exponential phase at 30°C in LB medium with 0.02% xylose. Cells were stained with membrane dye FM5-95 to visualize the contours of the cells and with DAPI to visualize the nucleoids. (A) Rok_{LS20}-IYPF (green) and membrane (red); (B) nucleoid (green) and membrane (red); (C) merged image of A and B with Rok_{LS20}-IYFP shown in green and nucleoid in red. Overlapping of the green and the red signals resulted in yellow. Distribution of Rok_{LS20}-IYFP was limited to the nucleoid in all cells analyzed (over 500 cells).

The images presented in Figure R7 show that the distribution of Rok_{LS20}-IYFP is limited to the region of the cell that is occupied by the nucleoid, the later visualized with 4'-6-diamidino-2-phenylindole (DAPI). This localization pattern is reminiscent of that observed for Rok_{Bs168} (120), suggesting that Rok_{LS20} may also have a genome-wide effect as recently demonstrated for Rok_{Bs168} (120). The observation that Rok_{LS20} co-localizes

with the nucleoid is consistent with its role as a transcriptional regulator.

4.2.7 Two distinct types of Rok proteins

Albano *et al.* (123) have reported the presence of Rok orthologs in genomes of two other *Bacillus* species closely related to *B. subtilis* (*B. licheniformis* and *B. amyloliquefaciens*), but not in other relatives such as *B. halodurans*, *B. anthracis* or *B. cereus*, and they proposed that *rok* was most likely introduced by horizontal transfer into an ancestor of a branch that contains *B. subtilis*, *B. licheniformis* and *B. amyloliquefaciens*. To gain insights into the evolutionary relationship between pLS20-encoded Rok_{LS20} with respect to other Rok proteins we carried out Blast and Psi-Blast searches using Rok_{LS20} as the query sequence. These searches led to the detection of 20 *rok* orthologs. The ClustalW2 program (http://www.ebi.ac.uk/clustalw) was then used to align the amino acid sequences of these Rok homologs, which were subsequently used to prepare distance-based tree dendrograms applying the neighbor-joining algorithm. These analyses revealed several interesting features. Firstly, except rok_{LS20} all other 19 *rok* genes identified are chromosome-located in the following Bacilli: *B. subtilis subsp subtilis* 168 (rok_{Bs168}); *B. subtilis subsp. spizizenii* strains W23, TU-B-10 and ATCC 6633; *B. subtilis* sp. BSn5; *Bacillus sp*. BT1B_CT2; *B. coagulans* strains 36D1 and 2-6; *B. amyloliquefaciens* strains FZB42 and DSM7; *B. pumilus* strains SAFR-032 and ATCC 14580; *B. atrophaeus* 1942; *B. licheniformis* ATCC 14580; and *B. pseudomycoides* DSM 12442. Secondly, based on several criteria, these 20 Rok orthologs can be divided into two groups, which we refer to as "large Rok proteins (group I)" and "small Rok proteins (group II)" (Fig. R8A). Group I constitutes eleven *rok* orthologs, which includes those present on *B. subtilis*, *B. licheniformis* and *B. amyloliquefaciens* identified previously by Albano *et al.* (2005). These eleven Rok proteins all have a similar relative large size ranging between 191 and 198 residues and they share an overall high level of identity (see Fig. R8B). In addition, a convergently oriented *yknT* gene flanks all these *rok* orthologs and all these *rok-yknt* modules are located in between the chromosomal genes *ykuV* and *mobA* (our unpublished results). The second group constitutes nine *rok* orthologs that includes rok_{LS20}. The Rok proteins encoded by these genes differ in various aspects from those of group I. Thus, (i) these Rok proteins are smaller and more heterogeneous in size (ranging between 132 and 186 residues), (ii) their overall level of identity is considerably lower than that observed for the other group of Rok proteins (compare Figs. R8B and R8C), and (iii) the genomic position of these *rok* genes in different species is not conserved.

At the beginning of this section it is mentioned that transformation levels were very low when cells contained pLS20cat. The results presented above show that *rok*$_{LS20}$ inhibits competence by repressing ComK expression. These results proved to be important for the rest of my PhD studies. I constructed a derivative of pLS20cat in which the *rok*$_{LS20}$ gene is inactivated (pLS20rok). Inactivation of *rok*$_{LS20}$ increased the competence level by a factor of about 30. However, as mentioned above (see table R1) competence levels of cells containing pLS20rok were still less than plasmid free cells. Therefore, I generally used a *B. subtilis* strain containing P$_{xyl}$-*comK* allowing the induction of high levels of competence also in the presence of pLS20cat.

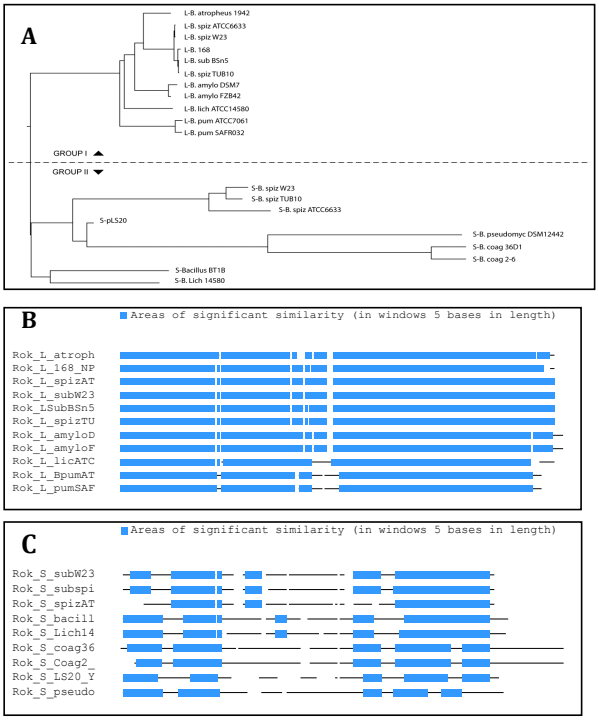

Figure R8. The set of twenty Rok proteins identified in a small number of related bacilli can be divided into two distinct groups. (**A**) Dendrogram of the probable relatedness of the twenty Rok proteins. The group I containing 11 "large" Rok proteins and Group II containing 9 "small" Rok proteins are indicated. (**B**) Alignment of the 11 large Rok proteins belonging to Group I. (**C**) Alignment of the 9 "small" Rok proteins belonging to Group II. Positions that contain identical or conserved changed residues are highlighted

4.3 Unraveling the regulatory circuit for conjugation of plasmid pLS20cat

The process of conjugation and its transcriptional regulation has been studied in considerable detail for various plasmids present in Gram-negative bacteria (10, 11, 41, 124). However, despite detailed studies on some enterococcal plasmids, relatively little is known about conjugation and especially its transcriptional regulation for plasmids of Gram-positive bacteria in general. A major objective of my thesis project was to advance our knowledge about the regulatory circuit for plasmid conjugation using pLS20cat as a model system for plasmids from Gram-positive bacteria.

4.3.1 pLS20cat conjugation is not activated by pheromones

Conjugation systems present on the Gram-positive *Enterococcus faecalis* plasmids pCF10, pAD1 and pPD1 are induced upon sensing a recipient-produced pheromone (61). To study whether conjugation of pLS20cat is also induced by pheromones we determined conjugation efficiencies in liquid medium under different conditions. Under the first condition, overnight grown cultures of donor (PKS11) and recipient (PKS7) cells were diluted and grown separately. At different times during growth, aliquots of the donor and recipient cells (~1:1 ratio) were mixed and their conjugation efficiencies were determined after a mating period of 15 min. The results presented in Figure R9 (broken lines) show that conjugation efficiencies increased during growth, reaching maximum levels near the end of the exponential growth phase, followed by a steep decrease in efficiencies at later times.

The increasing efficiency of conjugation observed during exponential growth phase might be explained by the accumulation of a conjugation activating signaling molecule produced by the recipient cells, as shown for *E. faecalis* plasmids (67). If this were the case, replacing the growth medium of the recipient cells with fresh medium before mixing with the donor cells should result in a reduction in conjugation efficiency (Fig R9, continuous lines).

We then considered the possibility that recipient cells were specifically competent for conjugation during the mid to late exponential growth phase. However this was not the case either as similar levels of conjugation efficiencies were obtained regardless of the growth stage of the recipient cells (in the range of 10^{-3}-10^{-4} transconjugants/donor). Altogether, these results indicate that the pLS20 conjugation system is not activated by recipient produced signaling molecules, but instead support the view that the conjugation system is continuously repressed except for a rather small window of time near the end of the exponential growth phase. This demonstrates that regulation of the conjugation system of pLS20 is fundamentally different from that of the enterococcal plasmids pCF10 and pAD1.

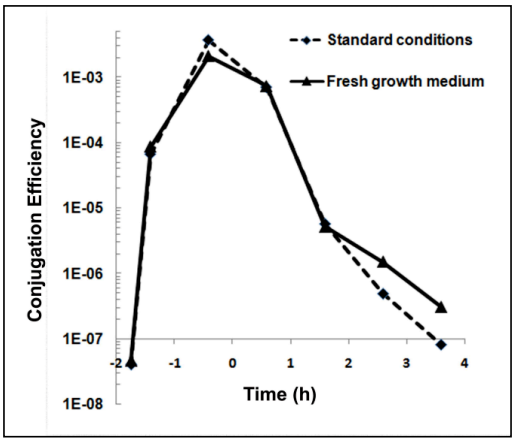

Figure R9. Conjugation kinetics of pLS20cat without and with replacing the growth medium. Overnight cultures of the donor and recipient cells (strain PKS11 and PKS7, respectively) were diluted in fresh pre-warmed LB medium to an OD_{600} of about 0.01 and grown separately at 37 ^{0}C. Growth was monitored by measuring OD_{600} at regular intervals. T=0 corresponds to the end of the exponential growth phase. At the indicated times, conjugation efficiencies were determined by mixing aliquots of donor and recipient cells and allowing a mating time of 15 min (broken line). In parallel experiments, recipient cells (continuous line) were pelleted and resuspended in equal volume of fresh LB medium before mating. Control experiments explain that centrifugation step did not affect conjugation.

4.3.2 Rco_{LS20}, an Xre-type regulator protein encoded by pLS20 gene 27c, represses conjugation of pLS20

The observation that efficient conjugation occurred only during a short time window raised the possibility that conjugation is kept in the default "OFF" state by a repressor, and switching to the "ON" state involves inactivation of the repressor. As presented above and shown in Figure R1, the large putative conjugation operon is flanked by a single divergently oriented ORF, ORF27c. The deduced protein sequence of ORF27c shows similarity to numerous bacterial transcriptional regulators, and based on homology putative protein encoded by ORF27c would belong to the Xre-type family of repressors. It is also predicted that putative protein p27c of pLS20cat contains a helix-turn-helix (HTH) DNA binding motif in its N-terminal region. An alignment of pLS20cat ORF27c with various Xre-type regulators is presented in Figure R10. This structural organization suggested that putative gene 27c encodes a repressor of the conjugation genes.

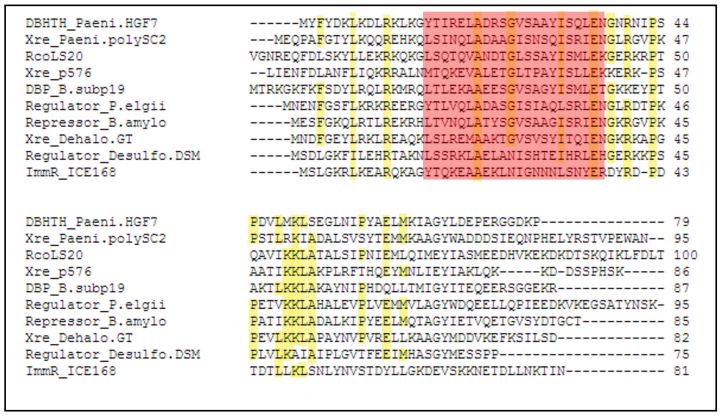

Figure R10. Alignment of different Xre-type repressors (Only N-terminal regions). Helix-Turn-Helix region is highlighted with red color. Highly conserved residues (at least 6 out of 10 proteins) are highlighted with yellow. Abbreviations (accession numbers given in brackets): DBHTH_Paeni.HGF7, DNA-binding helix-turn-helix protein of *Paenibacillus sp.* HGF7 (ZP_08510432); Xre_Paeni.polySC2, XRE family transcriptional regulator *Paenibacillus polymyxa* SC2 (YP_003945377); $Rco_{LS20,}$ Repressor of conjugation B. *subtilis* natto IFO 3335 plasmid pLS20 (YP_004243490); Xre_p576 , Xre type repressor *B.pumilus* NRS576 ; DBP_B.subp19, DNA binding protein of plasmid p19 *B.subtilis* 19 (ABP52080); Regulator_P.elgii, Transcriptional regulator *Paenibacillus elgii* B69 (ZP_09077606); Transcriptional Repressor_B.amylo, Transcriptional repressor RghR RapGH repressor *B. amyloliquefaciens* DSM 7 (YP_003921816); Xre_Dehalo.GT, XRE family transcriptional regulator *Dehalococcoides* sp. GT (YP_00346200); Regulator_Desulfo.DSM, putative transcriptional regulator *Desulfosporosinus youngiae* DSM 17734(ZP_09652311); ImmR_ICE168, XRE family transcriptional regulator of ICE element *B.subtilis*168 (NP_388363).

To test whether the putative gene 27c indeed encodes a repressor of conjugation, we studied if its ectopic expression affected conjugation. For this, we constructed strain PKS14 that harbours plasmid pLS20cat and contains a copy of gene 27c under the control of the IPTG-inducible P_{spank} promoter at the chromosomal *amyE* locus.

Table R3. Gene 27c (*rco*$_{LS20}$) encodes a repressor of conjugation

Strain	Genotype	Plasmid	IPTG (1mM)	Conjugation efficiency (transconjugants/donor)
PKS11	168 (wt)	pLS20cat	-	$5.6x10^{-3}$
			+	$3.8x10^{-3}$
PKS14	168,amyE::P_{spank}*rco*$_{LS20}$	pLS20cat	-	$1.6x10^{-4}$
			+	$<1x10^{-8}$
PKS86	168,amyE::P_{spank}*rco*$_{LS20}$	pLS20rco	-	$5.7x10^{-2}$
			+	$<1x10^{-8}$
Conjugation efficiencies are the mean value of at least three independent experiments				

Since maximum conjugation levels occur near the end of the exponential growth phase (see Fig. R9), we determined conjugation efficiencies of pLS20cat during this phase using PKS14 (in the presence/absence of IPTG). Strain PKS11 lacking the P_{spank}-27c construct was included as a control. The results presented in Table R3 show that ectopic

expression of gene 27c caused a dramatic decrease (>50,000 fold) in pLS20cat conjugation efficiency, supporting the view that it encodes a repressor of conjugation.

Based on these results and those presented below we denominated gene 27c of pLS20cat *rco*$_{LS20}$ (repressor of conjugation). To study the role of gene 27c in a more direct way, we constructed a derivative of pLS20cat, pLS20rco, in which gene 27c is deleted and replaced by a kanamycin resistance marker. To avoid possible constitutive de-repression of the conjugation operon that could pose a large burden to the cell, we introduced pLS20rco into strain PKS9 containing the P$_{spank}$-*rco*$_{LS20}$ construct. The resulting strain, PKS86, was used to determine the kinetics of conjugation during growth. Strain PKS14 (P$_{spank}$-*rok*$_{LS20}$, pLS20cat) was included as a control. Two interesting results were obtained for plasmid pLS20rco (see Fig. R11). First, conjugation efficiencies were higher at all time points measured; and second, conjugation levels were high for a very broad window of time. Therefore, in the absence of a functional *rco*$_{LS20}$ gene conjugation was no longer inhibited, most likely because the conjugation genes were not repressed.

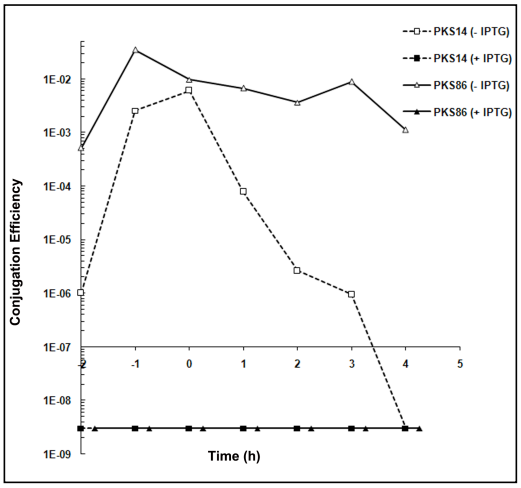

Figure R11. Conjugation kinetics of pLS20cat and pLS20rco were determined with and without ectopic expression of pLS20 gene 27c as described in Materials and Methods. PKS7 was used as recipient strain. The conjugation efficiencies obtained for strains PKS14 and PKS86 were below the detection level of 1x10^{-8} when grown in the presence of IPTG.ç

4.3.3 Transcriptional analysis of pLS20cat genes by RNAseq

Results shown above confirm that Rco$_{LS20}$ suppresses conjugation. To find out whether Rco$_{LS20}$ exerts its inhibitory effect on conjugation at the level of transcription and to identify genes that are under control of Rco$_{LS20}$, we performed RNAseq analysis. To do this, total RNA was isolated from late exponential growing cells (maximum conjugation state) of PKS11 (wt, pLS20cat) and of PKS14 (P$_{spank}$-*rco*$_{LS20}$, pLS20cat) grown in the presence of IPTG. In parallel, total RNA was isolated from plasmid-free *B. subtilis* 168 cells grown under same conditions to serve as a negative control. RNAseq experiments were carried out in a Facility Laboratory, namely Genomics Unit, Science

Park, Madrid, under the supervision of Dr. Ricardo Ramos Ruiz. Crude RNA sequences were analyzed by Ramon and David from Bioinformatics unit of our center (see M&M). Expression levels of pLS20cat genes were represented as a heat map (Fig. R12). The left lane of Figure R12 corresponds to the expression profile of pLS20cat genes when conjugation efficiency was at its maximum. Levels of expression are indicated by differentially colours of white to blue (white no expression and dark blue high expression). The middle lane shows the expression pattern in the presence of ectopic expression of *rco*$_{LS20}$.

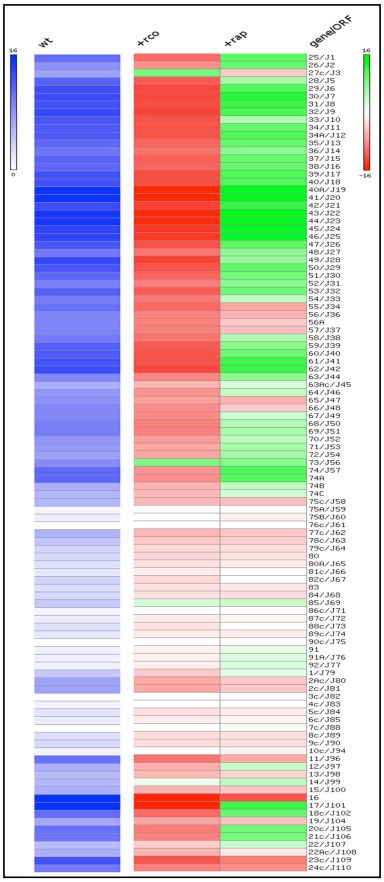

Figure R12. Heat map representation of the expression levels of the pLS20cat genes at late exponential phase under various conditions analysed by RNAseq. Left lane ("wt") shows the expression level of pLS20cat genes in the wild type strain background at late exponential phase when conjugation efficiency is at its maximum. Expression levels are presented on a $\log_2$ scale covering a range from 0 (white, lowest level) to 16 (blue, highest level). Middle (+Rco) and right (+Rap) lanes represent the effects of ectopic expression of Rco$_{LS20}$ (middle lane) or Rap$_{LS20}$ (right lane), respectively, on the expression of the pLS20cat genes. Differential expression levels are presented on a $\log_2$ scale covering a range of -16 to 16 using shades of red and green for repression and overexpression, respectively. White reflects no change in expression. Gene numbers according to our nomenclature and those deposited in database under accession number NC_015148.1 (preceded by "J") are given on the right). "c" corresponds to leftward oriented genes.

Increasing and decreasing RNA levels of individual genes are reflected by the intensity of green and red colors, respectively. The right lane (+rap) is explained below. Importantly, the heat map shows significantly reduced RNA levels for genes 28-72, as well as gene74, indicating that Rco$_{LS20}$ is responsible for repressing these genes. Some other genes outside the region spanning 28-74 are also repressed under these conditions. Thus, these results clearly show that Rco$_{LS20}$ represses almost all genes (28-74) located in the putative conjugation operon, as well as some other putative plasmid

genes encoding proteins of unknown functions. The combinations of the following results strongly indicate that Rco_{LS20} is the master regulator of conjugation. First, inactivation of rco_{LS20} on pLS20cat resulted in constitutive high levels of conjugation. Second, ectopic expression of rco_{LS20} resulted in undetectable conjugation level. And third, overproduction of Rco_{LS20} caused a strong repression of almost all genes in the putative conjugation operon.

4.3.4 Rco_{LS20} represses conjugation genes by regulating putative conjugation promoter

Taking into account the structural organization of pLS20cat genes 28 to 74, i.e. being all transcribed in the same direction, and the inhibitory effect that ectopic expression of rco_{LS20} has on expression of these genes, it is conceivable that expression of the conjugation operon is driven by a promoter located upstream of gene 28. To test this, we cloned the 570 bp intergenic rco_{LS20}-gene 28 region in the appropriate orientation in front of a promoter-less *lacZ* reporter and subsequently placed a single copy of this cassette at the *B. subtilis* chromosomal *thrC* locus, resulting in strain PKS3. Colonies of strain PKS3 were blue when grown overnight on LB agar plate supplemented with chromogenic substrate X-gal (Fig. R13). This result demonstrates that the rco_{LS20}-gene 28 intergenic region contains a promoter that we named P_c. Insights into the strength of the P_c promoter was obtained by performing the ß-gal assays of samples taken at different times during growth. The ß-gal activities obtained fluctuated between 300 and 500 Miller Units (MU) during exponential and stationary growth stages, respectively. This indicates that promoter P_c is a relative strong promoter that is probably not regulated by host-encoded factors when cells are grown under our experimental conditions.

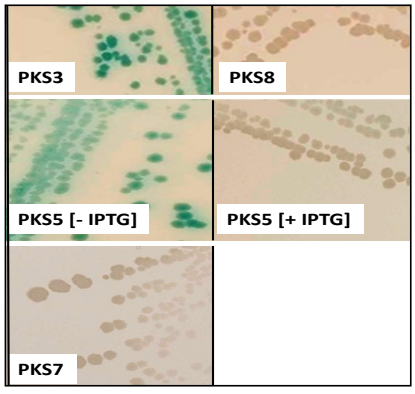

Figure R13. The rco_{LS20}-gene 28 intergenic region contains a strong promoter that is inhibited by the pLS20cat-encoded protein Rco_{LS20}. Strains were streaked on X-gal containing LB plates and incubated 16 hrs at 37^0 C. In case of PKS5, 10 µM IPTG was also added in the plate when indicated.

We next addressed the possibility that Rco_{LS20} would be directly responsible for suppressing conjugation by repressing the P_c promoter located upstream gene 28. For this we constructed strain PKS5 that contains, in addition to the P_c-*lacZ* fusion, a copy of rco_{LS20} under the control of IPTG inducible P_{spank} promoter and streaked this strain on LB plates with or without IPTG supplemented with X-gal. After overnight growth, colonies were blue when plates did not contain IPTG but they were white when plates contained 10 μM IPTG (Fig. R13). Although these results show that Rco_{LS20} represses promoter P_c when expressed from an inducible promoter, it does not guarantee that rco_{LS20} produces sufficient repressor to inhibit promoter P_c activity in its natural setting, i.e. on the plasmid. To study this we constructed strain PKS8 by introducing pLS20cat into strain PKS3 containing the P_c-*lacZ* fusion and plated it onto X-gal plates. Fig. R13 shows that colonies were white when cells contained pLS20cat. From these results, we concluded that Rco_{LS20} represses conjugation genes by regulating main conjugation promoter P_c. More detailed analyses about promoter P_c and its regulation by Rco_{LS20} is under investigation by another student in the lab.

4.3.5 Rap_{LS20} is not involved in sporulation or competence but stimulates conjugation

A putative *rap-phr* cassette (genes 25-26) is located downstream of the repressor gene rco_{LS20} in the plasmid pLS20cat; which we name rap_{LS20} and phr_{LS20}, respectively. The genome of *B. subtilis* contains eleven *rap* genes. The name *rap* refers to the activity of founding member RapA shown to be a Regulator Aspartate Phosphatase (125). Most Rap proteins studied interfere with developmental processes such as sporulation, competence and production of degradative enzymes and antibiotics (125–127). *rap* genes have also been identified on some plasmids from *B. subtilis* (97, 112, 128, 129). For those analyzed, plasmid-encoded rap genes also affect sporulation or production of extracellular proteases (129, 130). Based on this, it seemed plausible that rap_{LS20} too would play a role in sporulation and/or competence. To test this, we used strain GR20, which contains a single copy of rap_{LS20} gene under IPTG inducible promoter P_{spank} at its chromosome. Surprisingly though, overexpression of Rap_{LS20} did not significantly affect sporulation or competence (Table R5).

We then tested whether rap_{LS20} has a role in pLS20cat conjugation. We introduced pLS20cat into strain GR23, which contains the P_{spank}-rap_{LS20} fusion at the *amyE* locus. The resulting strain GR23 was used as a donor in conjugation experiments to determine the conjugation kinetics of pLS02cat in presence/absence of inducer. Strain PKS11 (wild type donor of pLS20cat) was included as a control (Fig. R14).

Interestingly, ectopic expression of *rap*$_{LS20}$ stimulated conjugation and the kinetics of conjugation obtained under these conditions was similar to that obtained for pLS20rco, the derivative containing a deletion of gene *rco*$_{LS20}$. Thus, in both cases, the maximum levels of conjugation increased and efficient conjugation occurred during a much broader time window.

Table R4. Ectopic expression of Rap$_{LS20}$ does not affect competence or sporulation

Process	IPTG (1mM)	Relative efficiency
Competence	-	1
	+	2.638
Sporulation	-	1
	+	0.65

Strain GR20 (*amyE*::P$_{spank}$-*rap*$_{LS20}$) was used to determine efficiencies of competence and sporulation using standard protocols.

To confirm that Rap$_{LS20}$ stimulates conjugation, we used two approaches. First, we determined the expression profile of pLS20cat genes by RNAseq when *rap*$_{LS20}$ was overexpressed from an ectopic locus in strain GR23. A heat map representation of the results is given in the right lane of Figure R12. Interestingly, almost all of the pLS20cat genes whose expression was repressed by Rco$_{LS20}$ (middle lane, red rectangles) were overexpressed when Rap$_{LS20}$ was induced ectopically (right lane, green rectangles). Second, we deleted rap$_{LS20}$ from pLS20cat by replacing it with a Kan marker, and then determined the conjugation kinetics of the resulting plasmid pLS20rap. Consistent with its role as a positive regulator, the absence of *rap*$_{LS20}$ resulted in a severe reduction in conjugation efficiency (PKS79) (Fig. R14).

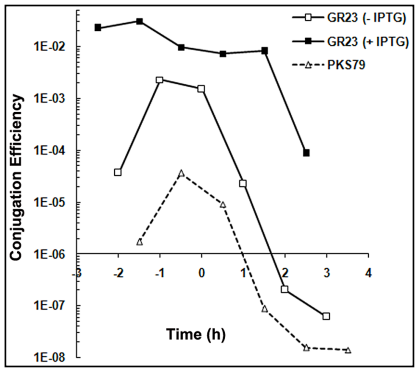

Figure R14. Rap$_{LS20}$ stimulates conjugation. Conjugation kinetics of pLS20cat and pLS20rap were determined with and without ectopic expression of Rap$_{LS20}$ as described in M&M. PKS7 was used as recipient strain. GR23 (pLS20cat) and PKS79 (pLS20rap) were used as donor strains. GR23 contains an ectopic copy of *rap*$_{LS20}$ under the control of the IPTG inducible P$_{spank}$ promoter at the chromosomal *amyE* locus. T=0 corresponds to the end of the exponential growth phase.

The combination of these results provides compelling evidence that Rap$_{LS20}$ stimulates conjugation by relieving Rco$_{LS20}$ mediated repression of the conjugation genes. Our preliminary data suggests that Rap$_{LS20}$ is directly responsible for inhibiting Rco$_{LS20}$ mediated repression of the P$_c$ promoter.

4.3.6 Phr*$_{LS20}$ determines the time window of efficient conjugation

Most *rap* genes are transcriptionally coupled to a downstream-located *phr* gene. The small *phr* genes encode a product that, after being subjected to an export-import-maturation process, produces a mature penta- or hexapeptide that inhibits the activity of its cognate Rap proteins. A putative *phr* gene, *phr*$_{LS20}$, is located immediately downstream of *rap*$_{LS20}$. The stop/start codons of these genes overlap and hence *phr*$_{LS20}$ is translationally coupled to *rap*$_{LS20}$, a similar situation occurs in many *rap-phr* cassettes. Inspection of the deduced protein sequence suggests that *phr*$_{LS20}$ indeed encodes a typical pre-pro-peptide. The 44-residue gene product is predicted to contain an N-terminal signal peptide, a conserved motif upstream of its predicted maturation cleavage site, as well as conserved residues within the putative mature peptide (97). Based on this, the mature *phr*$_{LS20}$-derived peptide is predicted to correspond to the five C-terminal residues of Phr$_{LS20}$, "QKGMY', which we will refer to as Phr*$_{LS20}$. We added chemically synthesized peptide to a culture of exponential growing cells and determined the conjugation efficiency of pLS20cat. Results presented in Figure R16A show that Phr*$_{LS20}$ greatly reduced the maximum level of conjugation. This supports the view that Phr*$_{LS20}$ inhibits Rap$_{LS20}$-mediated de-repression of the conjugation genes.

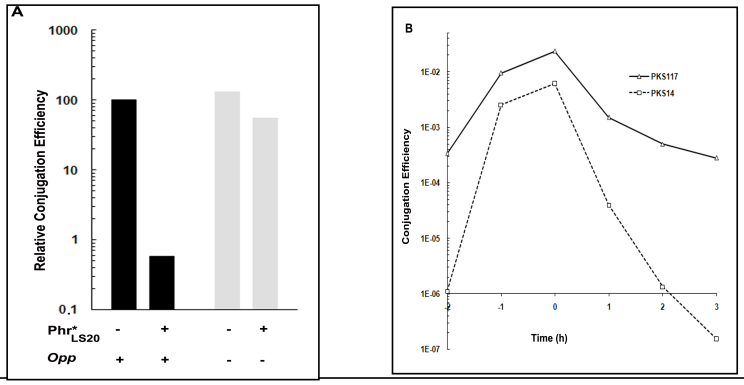

Figure R16. Phr*$_{LS20}$ pentapeptide inhibits conjugation in an *opp* dependent manner**. A.** Effects of synthetic Phr* peptide on conjugation in the wild type and an *opp* deficient background. Conjugation efficiencies of pLS20cat were determined at late exponential growth phase using as recipient strain PKS7, and as donor either strain PKS11 (wild type, black bars) or PKS98 (*oppA,* grey bars*).* Diluted overnight grown cultures of donor cells were split in two, and Phr*$_{LS20}$ pentapeptide was added to a final concentration of 6 μM to one of the cultures and equal volume of the peptide buffer to the other. **B.** Conjugation kinetics of pLS20cat and pLS20phr. Conjugation kinetics was determined as described in M & M using PKS7 as recipient strain and PKS14 (pLS20cat) or PKS117 (pLS20phr) as donor strains. T=0 corresponds to the end of the exponential growth phase. Both donor strains contain an ectopic copy of *rco$_{LS20}$* under the IPTG inducible P$_{spank}$ promoter at the chromosomal *amy*E locus. Overnight cultures of donor cells were grown in the presence of 1 mM IPTG and diluted in fresh pre-warmed LB medium without IPTG.

Mature Phr* peptides of other *rap-phr* cassettes are taken up by the oligo-peptide permease (Opp) of *B. subtilis* (47, 126, 127). Figure R16A shows that Phr*$_{LS20}$

also is taken up by Opp system as addition of Phr*$_{LS20}$ peptide hardly affected conjugation when donor cells were *opp*-deficient.

Next, we constructed a derivative of pLS20, pLS20phr, in which the *phr*$_{LS20}$ gene was deleted and tested its conjugation kinetics. The results presented in Figure R16B show that inactivation of phr$_{LS20}$ had similar effect on conjugation as those observed in the presence of ectopic expression of Rap$_{LS20}$ (Fig. R14) or inactivation of *rco*$_{LS20}$ (Fig. R11). Thus, in the absence of *phr*$_{LS20}$ conjugation efficiencies are high and conjugation occurs during a very broad time window.

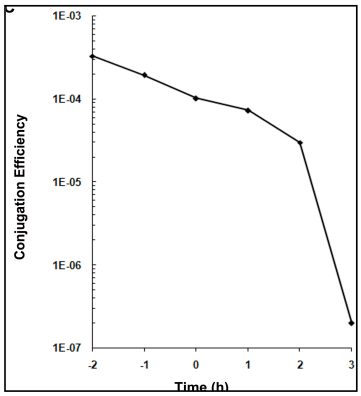

Figure R17. Conjugation kinetics of pLS20cat after re-dilution of the donor cell culture. Conjugation kinetics using PKS7 and PKS11 as recipient and donor strains, respectively, was determined as described in Materials and Methods with the following modification. Overnight cultures were diluted, grown until late exponential growth phase (OD_{600}= 0,8), and diluted again (to OD_{600}= 0.05) before starting the experiment.

Under our laboratory conditions, efficient conjugation is limited to a rather small time window before the end of the exponential growth phase (see Fig. R9). The results that Phr*$_{LS20}$ inhibits the activity of Rap$_{LS20}$, and that conjugation levels are high at all growth phases for pLS20phr indicate that the amount of Rap$_{LS20}$ protein is not a limiting factor for activating conjugation but that its activity is inhibited by Phr*$_{LS20}$ during early exponential as well as stationary growth phases. Phr*$_{LS20}$-mediated inhibition of conjugation during stationary phase is most likely due to the accumulation of Phr*$_{LS20}$ during growth, which will reach Rap$_{LS20}$-inhibiting threshold levels at or near the end of the exponential growth phase. However, the low levels of conjugation during early exponential growth cannot be explained by a similar kind of reasoning because the freshly diluted culture will contain low levels of Phr*$_{LS20}$ in the culture medium. One possible explanation for this is that this is due to intrinsic feature(s) of early exponential cells. This explanation can be excluded though as high levels of conjugation were obtained at early exponential growth phase with pLS20phr, pLS20rco or when Rap$_{LS20}$ was ectopically expressed. An alternative explanation could be that Rap$_{LS20}$-inhibiting levels of Phr*$_{LS20}$ are still present inside the cells after overnight grown culture were diluted in fresh medium. If this were the case, then high conjugation levels would be expected at early exponential growth phase by first growing the diluted overnight

culture of donor cells to the end of the exponential growth phase and then diluting it again. This experiment revealed that high conjugation levels were indeed observed at early exponential growth phase under these conditions (Fig. R17). Altogether, these results provide strong evidence that Phr*$_{LS20}$ is the determining factor in regulating the time window at which conjugation genes are activated. The observation that addition of synthetic Phr*$_{LS20}$ peptide inhibits conjugation suggests that the peptide acts in cell-cell signaling rather than being an autocrine signal.

5- Discussion

5.1 Rok$_{LS20}$ inhibits development of natural competency of *B. subtilis* 168

Conjugative plasmids not only confer a wide range of advantageous features to the host, but also can transmit these horizontally. While it is well known that plasmids often carry genes associated with resistance, virulence and degradative compound and bacteriocin production, genes regulating the natural competence of host cells have never been reported before. The discovery of a *rok* homologue in the genome of a natural conjugative plasmid was interesting as it provided a new dimension for studying plasmid-host interaction and evolution.

At present, we can only speculate about the biological function of pLS20-mediated inhibition of competence. The conjugation and competence processes have in common that they both transport single stranded DNA over the cell membrane and various similarities exist between the competence and conjugation related DNA transfer machines (131). It is possible that expression and assembly of the competence-related DNA uptake machine interferes with proper assembly and/or functioning of the conjugation-related DNA transfer machine. Rok$_{LS20}$-mediated inhibition of competence might suppress such a presumed interference. Whereas this is a plausible explanation it should be mentioned that, under laboratory conditions, maximum levels of conjugation and competence occur at different growth phases and conditions. It is also possible that the plasmid inhibits competence development to limit the fitness cost to the host. Alternatively, inhibition of competence may be beneficial for the genetic stability of plasmid pLS20 by (i) suppressing possible recombination between pLS20 absorbed exogenous DNA and/or (ii) suppressing possible inter- or intra-pLS20 recombination events. Regarding this latter supposition it is worth mentioning that pLS20 contains multiple regions of direct and inverted repeats.

The identification of a large number of new Rok homologues (Fig. R8) allowed us to compare and classify Rok proteins into two distinct groups. The difference in their genomic contexts and amino acid sequences between these two groups suggests that acquisition of these *rok* genes is the result of more than one independent horizontal transfer event. While the 'large *rok* genes' were most likely acquired through a single insertion event by a common ancestor, as suggested earlier (123), the origin of the 'small *rok* genes' is more complicated. It is possible that an ancient, common ancestor acquired a small *rok* gene long before the group of the 'large *rok* genes' appeared, and that enough time has passed for the gene to evolve independently in the different *Bacillus* species, which would explain why they are less conserved. Another possibility is that these 'small *rok* genes' were acquired independently throughout the history of

evolution. As one of the 'small *rok* genes' is located on the pLS20 plasmid, it is tempting to speculate that conjugative plasmids may have been involved in the spread of at least some of the *rok* genes belonging to this group. Interestingly, *B. licheniformis* ATCC 14580 and three *B. subtilis* subsp, each contain two *rok* genes: one of the group of 'large *rok* genes' and one of the groups of 'small *rok* genes'. No transformation has been observed for these four strains using standard protocols. However, it has been reported that artificial induction of *comK* expression in *B. subtilis* subsp. *spizizenii* strain ATCC 6633 is able to induce competence development (132). It will be interesting to test whether the lack of transformability in these strains is due to the presence of two *rok* genes on their genomes.

5.2 Phr*$_{LS20}$ peptide mediated regulation of conjugation system of pLS20

We report for the first time the regulation of a conjugation system present on a native *B. subtilis* plasmid. Our results show that the conjugation genes of pLS20cat are not induced by recipient-produced pheromones, demonstrating that regulation of the conjugation system of pLS20cat is fundamentally different from that of the enterococcal plasmids pAD1 and pCF10.

Using different experimental approaches we demonstrated that the pLS20cat gene 27c encodes the master regulator of conjugation, Rco$_{LS20}$. Interestingly, ectopic expression of Rco$_{LS20}$, predicted to be a DNA binding protein, resulted in the repression of not only the large, putative conjugation operon spanning genes 28 to 74, but also some other pLS20cat genes located outside the putative operon (for example, gene 11 and 16-21c). While it is possible that the effects of Rco$_{LS20}$ on the expression of some of the genes are indirect, the combination of our results clearly shows that Rco$_{LS20}$ is the master regulator of conjugation. Further work to characterize the DNA-binding properties of Rco$_{LS20}$ and to identify the operator site(s) of Rco$_{LS20}$ will be able to provide important information on how the different genes on pLS20 are regulated.

We also show that conjugation is activated by anti-repression and that Rap$_{LS20}$, encoded by pLS20cat gene 25, is the anti-repressor of Rco$_{LS20}$. Rap$_{LS20}$ belongs to the large family of Rap proteins. At the moment, the number of *rap* genes present in databases exceeded 500 members. To our knowledge, this is the first time that a Rap protein has been demonstrated to activate plasmid conjugation.

The observations that Rap$_{LS20}$ did not affect these differentiation routes may be explained by the recently obtained functional and structural data on how Rap proteins interact with regulatory proteins in the sporulation and competence pathways (133, 134). Initiation of sporulation is controlled by the master regulator of sporulation,

Spo0A, which becomes activated upon phosphorylation through a phosphorelay. Eight of the Rap proteins encoded by the chromosome of *B. subtilis* and some Rap proteins encoded by the *Bacillus* plasmids have been shown to interact with and dephosphorylate the sporulation protein Spo0F, one of the intermediate signal transducers. This interrupts the phosphate flux in which the phosphate is transferred from kinases to Spo0A through a phosphorelay (135). Competence development, on the other hand, is controlled by the transcription factor of competence, ComA. It has been shown that RapC, RapF and RapH inhibit competence by interacting with ComA and preventing it from binding to DNA (136). Sequence comparison show that all Rap proteins contain a rather small N-terminal domain of about 70 residues that is composed of a 3-helix bundle. This N-terminal domain is connected by a flexible linker to a much larger C-terminal domain that generally harbors the Rap characteristic tetratricopeptide repeat (TPR) sequences (133, 134). The recently resolved crystal structure of the Spo0F-RapH complex revealed that Spo0F interacts with both the C-terminal TPR domain and the N-terminal 3-helix bundle of RapH, including Gln47 in the N-terminal domain. This glutamine residue (GLu49 in the case of RapP encoded by the *B. subtilis* plasmid pBS32) is highly conserved and constitutes the catalytic residue responsible for dephosphorylating Spo0F~P (47, 130, 137). The alignment of the N-terminal regions of Rap proteins, presented in Figure R15, shows that neither the catalytic residue nor other residues in this region shown to be important for RapH phosphatase activity *in vitro* and *in vivo* are conserved in Rap_{LS20} or Rap_{576}, the latter is encoded by a related theta replicating plasmid p576 (112).

In the case of ComA, several ComA-interacting residues of RapF, which are conserved among Rap proteins known to interact with ComA, have been identified and shown to be vital for the functionality of RapF (134). The alignment in Figure R15 shows that these residues are not conserved in Rap_{LS20} or Rap_{576}, consistent with our finding that Rap_{LS20} does not affect competence. Thus, residues important for interaction with Spo0F or ComA are not conserved in Rap_{L20}, which most probably explains why Rap_{LS20} does not affect sporulation or competence. It is worth mentioning that Rap proteins involved in the regulation of competence and sporulation pathways act as modulators, by inhibiting and/or delaying these developmental processes. On contrary, Rap_{LS20} functions as an activator, and rather than being a modulator, it plays a decisive role in the conjugation process by relieving Rco_{LS20}-mediated repression. Our results on Rap_{LS20}, together with published results on other Rap proteins demonstrates the enormous plasticity of how these proteins have evolved into versatile regulatory proteins that

control diverse cellular processes by interacting with a wide range of regulatory proteins.

Figure R15. Conserved residues important for Rap proteins known to interact with Spo0F or ComA are not conserved in Rap$_{LS20}$. Amino acid sequences corresponding to the N-terminal 3-helix bundles of *Bacillus* Rap proteins previously demonstrated to dephosphorylate Spo0F (RapP, RapA, RapB, RapE, RapJ RapH, Rap$_{XO1}$ (=BXA0205), and Rap60 [Spo0F-phosphatase activity has not been demonstrated biochemically for Rap60]) or to interact with ComA (RapF, RapC and RapH) were aligned together with the corresponding regions of Rap$_{LS20}$, Rap$_{576}$ and RapI using clustallW2. The highly conserved tryptophan residue present in all these Rap proteins is indicated in green. The catalytic Gln47 residue of RapH that is conserved in six of the seven other Spo0F-interacting Rap proteins as well as in RapI are highlighted in red. Alanine substitutions in Rap proteins that cause complete or significant loss of function/interaction with Spo0F and ComA are highlighted by blue boxes (133)(134). RapH residue Leu55 is conserved in Rap$_{LS20}$ and Rap$_{576}$. It is worth mentioning that although the L55A mutant affected the function of RapH *in vivo*, no loss of RapH function was observed for this mutant *in vitro* (133). Positions of the α-helices are indicated above the alignment.

The combination of results obtained indicate that the pLS20 conjugation genes will be activated specifically when donor cells are surrounded by recipient cells (when Phr*$_{LS20}$ concentration is low), and not by other donor cells (when Phr*$_{LS20}$ concentration is high). Besides preventing futile expression of conjugation genes when recipient cells are not present, strict regulation of the conjugation genes is likely to serve other purposes. For instance, pLS20 replicates via a theta mode of replication (97). During conjugation, however, replication switches to the rolling-circle mode to generate the ssDNA strand that is destined to be transferred into the recipient cell. Simultaneous replication of the theta and rolling-circle modes are likely not compatible and strict regulation of the conjugation genes, amongst which are those involved in initiating rolling circle replication, contributes to selecting the mode of replication according to the situation. In addition, it is likely that expression of the conjugation genes poses a large burden to the cell. This view is supported by our observations that growth is affected in cells harbouring pLS20rco and pLS20phr, i.e. plasmids containing alterations leading to constitutive expression of the conjugation genes.

A summary of the regulatory circuitry of the pLS20 conjugation genes is schematically presented in Figure R18. Rco$_{LS20}$ is responsible for maintaining conjugation in the default "OFF" state by repressing the conjugation genes. Rap$_{LS20}$ can

activate conjugation by relieving Rco_{LS20}-mediated repression, but is only able to do so when its activity is not inhibited by the Phr^*_{LS20} signaling peptide. Therefore, conjugation of the pLS20cat plasmid is strictly regulated by the Phr^*_{LS20} peptide-mediated quorum sensing (QS) mechanism.

QS is a common way by which bacteria communicate with one another using small and diffusible chemical signaling molecules. When the concentration of a signaling molecule reaches a certain "quorum", bacteria respond by altering its gene expression profile at a (sub) population-wide scale. Several cellular processes in both Gram-positive and Gram-negative bacteria have been shown to be regulated by QS, among them the development of natural competence in *B. subtilis* and *Streptococcus pneumonia* (126, 138). Here, we show that QS plays a role in HGT at another level by regulating expression of conjugation genes of plasmid pLS20. So far, QS has been reported to regulate conjugation genes of only a few other conjugative elements. These include the transfer of the tumor-inducing pTI plasmid of the Gram-negative *Agrobacterium tumefaciens* into plant cells. In this case, activation of conjugation requires two signaling peptides, one produced by the plant and the other by the donor cell (139). Regulation of conjugation of the enterococcal plasmid pCF10, and probably in a similar way pAD1, also involves two signaling peptides, one produced by donor and the other by recipient cells. The two peptides compete for binding to a single transcriptional regulator, PrgX, and act antagonistically on conjugation. However, instead of being an activator, PrgX is a repressor. When PrgX is bound to the donor-produced signaling peptide the complex binds DNA and represses the conjugation genes. Conjugation genes become activated when recipient-produced signaling peptide replaces the donor-produced signaling peptide in the PrgX/peptide complex thereby inactivating the repressor activity of PrgX. Consequently, conjugation genes are activated by recipient produced signaling peptides (67).

Our results show that the QS mechanism to regulate conjugation genes of pLS20 differs in various aspects from those regulating conjugation of the pTi and pCF10/pAD1 plasmids. First, regulation of pLS20 conjugation genes involves not two but only one signaling peptide, Phr^*_{LS20}. Second, the signaling peptide does not act directly on the transcriptional regulator but instead regulates activity of another protein, Rap_{LS20}, which functions as an anti-repressor. And third, the signaling peptide does not function to activate conjugation genes but to return the conjugation system to the default "OFF" state by inhibiting the activity of the anti-repressor.

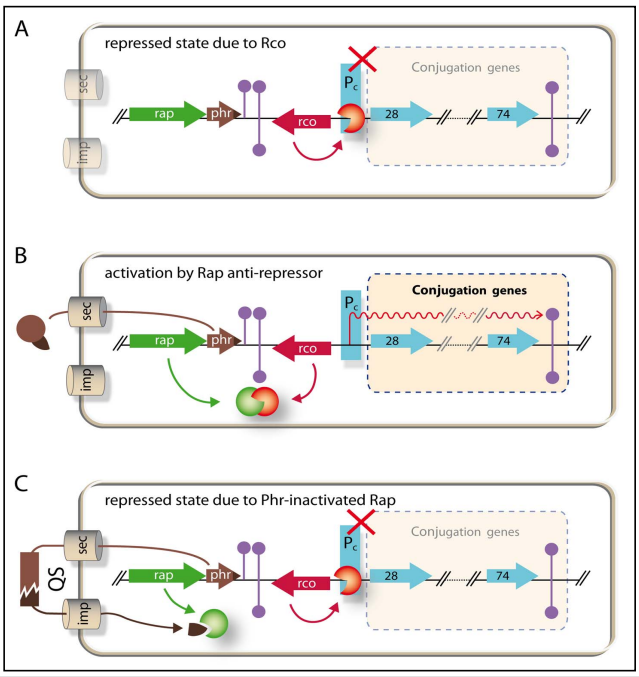

Figure R18. Model of regulatory circuitry of pLS20 conjugation genes. A. Repressed state due to Rco_{LS20}. Gene *rco*$_{LS20}$ (red arrow, rco) encoding the master repressor of conjugation genes Rco_{LS20} is divergently transcribed from the conjugation operon encompassing genes 28 to 74 (light blue arrows). Rco_{LS20} inhibits expression of the conjugation genes by repressing the promoter, P_C, located upstream of gene 28, the first gene of the conjugation operon. **B.** Activation of conjugation by Rap_{LS20} anti-repressor. Gene *rap*$_{LS20}$ (green arrow, rap) encodes the anti-repressor of Rco_{LS20} leading to de-repression of the conjugation genes. **C**. Repressed state due to inactivation of Rap_{LS20} by signaling peptide Phr^*_{LS20}. Gene phr_{LS20} (brown arrow, phr) encodes a pre-pro protein of 44 residue. This protein is subject to an export-maturation-import route. The mature pentapeptide inhibits activity of the Rap_{LS20} anti-repressor protein. For simplicity, import of the mature peptide has been shown in the peptide-producing cell. Grey cylinders labeled sec and imp indicate the secretion and import routes, respectively. Extracellular processing of the secreted peptide is symbolized by the brown interrupted rectangle. QS, quorum sensing.

Although *rap-phr* cassettes have not been shown before to regulate conjugation of a plasmid, the *B. subtilis* chromosomal *rapI-phrI* cassette has been described to regulate activation of the integrative conjugative element ICEBs1 (47). There are several similarities but also interesting differences between the conjugation systems present on plasmid pLS20 and the chromosomal ICEBs1. In both systems, transcription of the conjugation genes are repressed by an Xre-type repressor (Rco_{LS20} and ImmR, respectively) and the gene encoding the repressor protein is divergently oriented with respect to a large putative operon encoding the structural conjugation genes. In addition, in both cases conjugation is activated by a Rap protein (Rap_{LS20} and RapI,

respectively) whose activity is controlled by a quorum sensing peptide encoded by the downstream *phr* gene. As we have proposed for rap_{LS20}-phr_{LS20}, a major function of the *rap*I-*phr*I cassette of ICEBs1 is a sensing mechanism to induce genes required for conjugation when recipient cells are present (47).

However, there are several important differences between the Rap_{LS20}-Phr_{LS20} and the RapI-PhrI systems. One significant difference is that transfer of ICEBs1 requires, besides RapI, the ICEBs1-encoded ImmA protein to relieve the ImmR-mediated repression of the ICEBs1 genes (140). The *immA* gene is located immediately downstream of *immR*. ImmA is a protease that cleaves ImmR and its activity is probably controlled by RapI (140). However, an immA homologue is not present on pLS20cat. More importantly, our preliminary results indicate that Rap_{LS20} functions directly as the anti-repressor of Rco_{LS20} (to be published elsewhere). Another major difference is that efficient mating of ICEBs1, like most other ICEs, is limited to solid media, whereas pLS20 mates also efficiently in liquid medium (13, 14). In a typical solid medium conjugation experiment, a concentrated mixture of donor and recipient cells is incubated on a solid surface for several hours, permitting donor cells to intimately contact recipient cells for a prolonged period of time. These conditions also correlate with high cell density, cell crowding and starvation, i.e. stationary phase conditions at which cells generally induce developmental pathways in order to cope with the suboptimal growth conditions.

The third difference is the timing of transfer. Auchtung et al. (47) reported that conjugation of ICEBs1 is low during mid exponential and much higher during stationary growth. The authors provided evidence that *rapI* is negatively regulated by the transition state regulator AbrB, which represses transcription of several *B. subtilis* genes specifically during the exponential growth phase(141). Thus, RapI stimulates transfer of ICEBs1 during stationary phase conditions that favor intimate intercellular contacts; conditions that typically occur during conjugation on solid medium. In *B. subtilis* such conditions also stimulate initiation of the sporulation pathway. Most probably, sporulation and conjugative transfer of an ICE element are not compatible and hence efficient transfer of the ICEBs1 element would benefit from inhibiting sporulation in that cell. Interestingly, we have found that ectopic expression of RapI strongly affects sporulation. In agreement with our results, it has been shown very recently that RapI can dephosphorylate Spo0F *in vitro* (137). Together these results demonstrated that RapI has a dual function: it activates transfer of ICEBs1 during stationary phase and inhibits the initiation of sporulation that is normally stimulated under these conditions.

We have shown that Rap_{LS20} regulates conjugation of pLS20cat in a strikingly different manner. Several results showed that efficient pLS20cat conjugation occurs during exponential growth and that it is strongly inhibited during stationary growth. This important difference may be related to the fact that conjugation of pLS20cat occurs efficiently in liquid medium when cells have a planktonic lifestyle and probably spend more time in the exponential growth phase than cells growing in sessile communities. Our results on Rap_{LS20}, together with published results on other Rap proteins, demonstrate the enormous plasticity of how these proteins have evolved into versatile regulatory proteins that control diverse cellular processes by interacting with a wide range of other regulatory proteins.

6- Conclusions

1. Plasmid pLS20cat has a size of 65,774 bps and its (G+C) content is 37.8%. It contains 92 ORFs that would encode for a protein of at least 30 amino acids, and that are preceded by a putative ribosomal binding site. Based on annotation and other criteria, the genes/ORFs have been grouped in five modules.
2. Presence of plasmid pLS20cat inhibits competency of its host. ORF64 of pLS20cat is responsible for competence inhibition. The deduced amino acid of ORF64 has homology to Rok_{Bs168} encoded by the genome of *Bacillus subtilis* and thus we named the gene rok_{LS20}.
3. Repression of *comK* expression by preferential binding of Rok_{LS20} to the *comK* promoter results in decreased numbers within a population that become competent.
4. Rok_{LS20} does not affect ComK levels posttranscriptionally. Hence, a strain allowing ectopic expression of ComK from an inducible promoter develops high levels of competence even in the presence of plasmid pLS20cat.
5. pLS20cat conjugates in liquid media and maximum conjugation occurs during a short window of time near the end of the exponential growth phase.
6. The conjugation genes of pLS20cat are not induced by recipient-encoded pheromones and thus regulation of the conjugation system of pLS20cat is fundamentally different from that of the enterococcal plasmids pAD1 and pCF10.
7. A long conjugation operon followed by *rap-phr* cassette is present next to origin of replication.
8. The deduced protein sequence of ORF27c shows homology to Xre-type repressors. Ectopic overexpression of gene 27c fully suppresses pLS20cat conjugation. We named it rco_{LS20} (repressor of conjugation). Binding of Rco_{LS20} at promoter P_c, located upstream of gene 28, the first gene of conjugation operon, represses expression of the genes present in the conjugation operon.
9. Rap_{LS20} does not play any role in sporulation or competence rather it activates conjugation by counteracting the effect of Rco_{LS20}.
10. Mature pentapeptide Phr^*_{LS20} decides the timing of conjugation by regulating the activity of Rap_{LS20}.

6-Conclusiones

1. El plásmido pLS20cat tiene un tamaño de 65,774 pb y su contenido en G+C es del 37,8%. Contiene 92 ORFs que pueden codificar proteínas de al menos 30 aminoácidos que están precedidos de un RBS putativo. Basándose en la anotación y otros criterios, los genes/ORFs han sido agrupados en cinco módulos.
2. La presencia del plásmido pLS20cat inhibe la competencia del hospedador. El ORF64, denominado *rok*$_{LS20}$, de pLS20cat es el responsable de la inhibición de la competencia.
3. La represión de la expresión de *comK* por la unión preferencial de Rok$_{LS20}$ al promotor de *comK* produce un descenso en el número de células de una población que entran en competencia.
4. Rok$_{LS20}$ no afecta a los niveles de ComK de forma postranscripcional. Por lo tanto, una cepa que permite la expresión ectópica de ComK mediante un promotor inducible, desarrolla altos niveles de competencia incluso en presencia del plásmido pLS20cat.
5. El plásmido pLS20cat puede conjugar en medio líquido y la mayor eficiencia de conjugación ocurre durante un corto periodo de tiempo cercano al final de la fase exponencial de crecimiento.
6. Los genes de conjugación de pLS20cat no son inducidos mediante feromonas codificadas por el receptor, y por lo tanto la regulación del sistema de conjugación de pLS20cat es diferente a la encontrada los plásmidos pAD1 Y pCF10 de enterococos.
7. Junto al origen de replicación se encuentra un operón de conjugación seguido de un casete *rap-phr*.
8. La proteína deducida de la secuencia del ORF27c muestra homología con represores tipo Xre. La expresión ectópica del gen 27c reprime totalmente la conjugación de pLS20cat. Hemos nombrado a este gen *rco*$_{LS20}$ (represor de la conjugación). La unión de Rco$_{LS20}$ al promotor P$_c$, situado aguas arriba del gen 28, el primer gen del operón de la conjugación, reprime la expresión de los genes presentes en el operón de conjugación.
9. Rap$_{LS20}$ no juega ningún papel en la esporulación o la competencia, sin embargo activa la conjugación contrarrestando el efecto de Rco$_{LS20}$.

10. El pentapéptido maduro Phr^*_{LS20} decide el tiempo de conjugación mediante la regulación de la actividad de Rap_{LS20}.

7- Presentación

Presentación

Introducción

Las bacterias pueden intercambiar genes de resistencia a antibióticos mediante diferentes procesos, conocidos conjuntamente como Transferencia Horizontal de Genes (horizontal gene transfer, HGT). La dispersión entre bacterias patógenas de los genes responsables de las resistencias a antibióticos, es un gran problema tanto para la salud humana como en el ámbito veterinario. La HGT puede producirse mediante tres principales mecanismos, los cuales son 1) transformación, 2) transducción y 3) conjugación. La transformación consiste en la absorción del DNA desnudo exógeno por células competentes, integrándolo en el genoma mediante recombinación homóloga. La transducción es la transferencia de genes bacterianos de una célula a otra mediante fagos. Finalmente, la conjugación es un proceso altamente específico por el cual el DNA es transferido desde una célula donadora a otra receptora mediante un complejo multiproteico especializado, denominado aparato de conjugación. Al contrario que la competencia y la transducción, la conjugación requiere un contacto directo entre las células donadoras y receptoras. La conjugación está muy extendida entre bacterias Gram-negativas y Gram-positivas y está mediada por sistemas de conjugación que pueden estar presentes en plásmidos y/o en elementos de integración conjugativos (ICE, Integrative Conjugative Elements).

Los plásmidos juegan un importante papel en la HGT y también son herramientas esenciales en la ingeniería genética. Los plásmidos codifican genes relacionados con la detoxicación, virulencia, interacciones ambientales, y como se ha mencionado anteriormente, resistencia a antibióticos. Por lo tanto, para poder entender estos importantes rasgos bacterianos, a menudo implicados en la salud humana o el bienestar, es esencial conocer como funcionan los plásmidos conjugativos.

Generalmente, la expresión de los genes del plásmido necesarios para su transferencia por conjugación está estrictamente regulada para minimizar la carga sobre la célula. Estos sistemas permanecen en un estado apagado y su transferencia es inducida en respuesta a señales ambientales y fisiológicas, así como la presencia de células receptoras. La comprensión de estos sistemas de control pueden ayudar a modificar la gestión de las comunidades microbianas donde la transferencia de plásmidos es deseable o indeseable.

El proceso de conjugación y su regulación transcripcional ha sido estudiada en considerable detalle en varios plásmidos presentes en bacterias Gram-negativas. Sin embargo, se sabe relativamente poco sobre los sistemas de conjugación en plásmidos

precedentes de bacterias Gram-positivas, muchos de los cuales son organismos industrialmente o medicamente importantes. Para aumentar el conocimiento sobre sistemas de conjugación y su regulación en bacterias Gram-positivas, estudiamos el plásmido nativo pLS20 de *Bacillus subtilis*. Los estudios descritos en esta tesis implican (i) la secuenciación y la anotación del plásmido pLS20cat (ii) la caracterización de un gen de pLS20 responsable de la inhibición de la ruta de la competencia en la célula portadora, y (iii) el análisis del circuito de la regulación del proceso de conjugación.

Objetivos

Hemos elegido el plásmido pLS20cat para estudiar los mecanismos que regulan de la conjugación. En esta tesis se proponen los siguientes objetivos.

1. Secuenciación y anotación del plásmido pLS20cat.
2. Caracterización de los factores responsables de la inhibición de la competencia en presencia de pLS20cat.
3. Determinación de la dinámica de conjugación de pLS20cat.
4. Caracterización de las rutas necesarias para regular la expresión de los genes de conjugación en pLS20cat.

Resultados y discusión

El primer objetivo era secuenciar y anotar todo el plásmido pLS20cat. La secuenciación del plásmido pLS20cat reveló que posee un tamaño de 65.774 pares de bases (pb) y que su contenido en G+C es del 37,8%. Se han identificado un total de 92 marcos de lectura abiertos (ORFs, Open Reading Frames) de más de 30 codones. Un resumen de las características de los genes/ORFs identificados, incluyendo sus posible lugares de unión del ribosoma (RBS, Ribosomal Binding Site) del plásmido pLS20cat se muestran el la Tabla Suplementaria S1.

Las características identificadas, combinadas con la información del plásmido relacionado p576, nos ha permitido clasificar las regiones y los diferentes genes/ORFs de pLS20cat en cinco módulos, (i) funciones de partición del plásmido, (ii) origen de replicación, (iii) casete *rap-phr*, (iv) posible operón de transferencia por conjugación con un gen contiguo que codifica un posible represor, y (v) genes/ORFs de funciones desconocidas. Estos cinco módulos están indicados en diferentes colores en el mapa esquemático de pLS20cat mostrado en la figura R1.

El segundo objetivo principal era determinar el mecanismo por el cual pLS20cat inhibe el desarrollo de la ruta de competencia de su hospedador. La proteína deducida de la secuencia del ORF64 de pLS20cat muestra similitud con Rok_{Bs168} codificada en el genoma de *B. subtilis*. Rok_{Bs168} limita la subpoblación de células que entran en competencia en un cultivo reprimiendo *comK*, que codifica el principal regulador de los genes de competencia. Basándose en la homología identificada, denominamos al ORF64, rok_{LS20}. Hemos demostrado que la sobreexpresión de rok_{LS20} mediante un locus ectópico inhibe la competencia. La activación de la expresión de varios genes de competencia, incluyendo *comG*, necesitan niveles de ComK altos. Usando una fusión transcripcional *comG-gfp*, hemos determinado el efecto de la expresión de rok_{LS20} sobre el porcentaje de células de un cultivo que desarrollan competencia. La sobreexpresión de rok_{LS20} causa una disminución en el número de células que expresan GFP, lo que implica que menos células expresan niveles de ComK lo suficientemente altos como para iniciar el desarrollo de competencia. Estos resultados indican que Rok_{LS20} afecta a la cantidad funcional de ComK en la célula. Hemos demostrado que Rok_{LS20} se une preferentemente al promotor de *comK* y que no afecta a la competencia cuando *comK* se encuentra bajo el control de un promotor inducible. Estos resultados indican que Rok_{LS20} regula la cantidad de ComK a nivel transcripcional pero no a nivel post-transcripcional.

Una explicación atractiva de porqué pLS20 contiene un gen que inhibe la competencia es la que se da a continuación. Durante la conjugación, el DNA de pLS20 es transferido desde la célula donadora hasta la célula receptora. El DNA es también transferido durante la competencia; en este caso el DNA se recoge del ambiente circundante. En ambos casos, la transferencia de DNA implica un elaborado complejo multiproteico localizado en la membrana. Posiblemente, la expresión y el ensamblaje de la maquinaria de la captación de DNA relacionadas con la competencia, interfiere con el ensamblaje y/o el funcionamiento de los mecanismos de transferencia de DNA relacionados con una adecuada conjugación. La inhibición de la competencia mediada por Rok_{LS20} podría suprimir una presunta interferencia de este tipo.

La tercera parte de esta tesis describe la caracterización del circuito regulatorio presente en el plásmido pLS20. La cinética de conjugación de pLS20cat muestra que no es activada por feromonas que producen las células receptoras. Por lo tanto, la regulación del proceso de conjugación es fundamentalmente diferente de la de los plásmidos de enterococos. Hemos observado que la conjugación se realiza de forma eficiente durante un corto periodo de tiempo, lo que planteó la posibilidad de que la conjugación se mantiene en un estado inactivo por defecto, mediante una proteína represora de la transcripción y que se activa sólo en un cierto periodo durante el crecimiento mediante una inactivación

temporal del represor del proceso de conjugación. El gen putativo 27c de pLS20cat flanquea un gran operón putativo de conjugación. Debido a que la secuencia prevista de la proteína del gen 27c muestra homología con reguladores transcripcionales tipo Xre, éste gen era un buen candidato para participar en la regulación de los genes de conjugación. La sobreexpresión ectópica del gen 27c inhibió completamente la conjugación, mientras que la delección de este gen en pLS20cat aumentó la eficiencia de conjugación durante todas las fases del crecimiento. Se realizó un análisis del transcriptoma de pLS20cat mediante el método de RNAseq. Este análisis mostró una reducción de los niveles de expresión de todos los genes presentes en el operón putativo de la conjugación cuando el gen 27c se expresa a partir de un locus ectópico. Sobre la base de estos resultados, denominamos al gen 27c como *rco*$_{LS20}$ (represor de la conjugación). Los resultados presentados en esta tesis demuestran que *rco*$_{LS20}$ suprime la expresión de los genes de la conjugación mediante la represión del principal promotor de la conjugación, P_c, localizado "upstream" del primer gen del operón de conjugación, el gen 28.

Situado aguas abajo de *rco*$_{LS20}$ se encuentra un casete putativo *rap-phr* (genes 25-26); a los que llamamos *rap*$_{LS20}$ y *phr*$_{LS20}$, respectivamente. Otros casetes *rap-phr* están presentes en los genomas de muchos bacilos, así como en sus plásmidos. La mayoría de las proteínas Rap estudiadas desempeñan un papel en el desarrollo de las rutas de la competencia y la esporulación. Curiosamente, la expresión ectópica de Rap$_{LS20}$ no afecta significativamente a la esporulación o a la competencia. Por el contrario, hemos encontrado que la expresión ectópica de Rap$_{LS20}$ aumenta la eficiencia de la conjugación de pLS20cat. El análisis mediante RNAseq demostró que la expresión ectópica de Rap$_{LS20}$ provoca una sobreexpresión de casi todos los genes cuya expresión fue reprimida por Rco$_{LS20}$. Por lo que conocemos, esta es la primera vez que se ha demostrado que una proteína Rap puede activar la conjugación de plásmidos. Los análisis estructurales y genéticos han identificado que la región N-terminal de varias proteínas Rap interactúan con las proteínas Spo0F y ComA, implicadas en la activación de la esporulación o la competencia, respectivamente, y se han identificado varios residuos cruciales conservados para estas interacciones. La comparación de las regiones N-terminales de las proteínas Rap muestra que ni los residuos importantes para la interacción con Spo0F ni ComA se conservan en Rap$_{LS20}$, lo que probablemente explique por qué Rap$_{LS20}$ no afecta a la esporulación o la competencia.

La mayor parte de los genes de *phr* están acoplados transcripcionalmente con un gen *phr* situado upstream. Los pequeños genes *phr* codifican un producto que, después de ser sometido a un proceso de exportación, maduración e importación, inhibe la actividad de sus proteínas Rap afines. Hemos probado que *phr*$_{LS20}$ tiene un papel en la regulación de

la actividad de Rap_{LS20}. La adición del péptido maduro Phr_{LS20} sintetizado químicamente (Phr^*_{LS20}, corresponde a los cinco residuos C-terminales de Phr_{LS20}), bloquea la conjugación, y la inactivación de phr_{LS20} en el plásmido hace que se destruya la regulación de la conjugación de forma temporal. En concreto, la inactivación de phr_{LS20} tiene efectos similares a la expresión ectópica de rap_{LS20}.

En conjunto, los efectos ejercidos por Phr^*_{LS20} explican la gran disminución en la eficiencia de la conjugación de pLS20cat cuando las células alcanzan la fase estacionaria del crecimiento, y encajan con la visión de que la conjugación se desactiva cuando la mayoría de las células contienen el plásmido y, por lo tanto, evitar intentos inútiles de conjugar en ausencia de células receptoras. Sin embargo los resultados no explican por qué los niveles de conjugación de pLS20cat son muy bajos durante la primera fase del crecimiento exponencial. Consideramos en primer lugar, la posibilidad de que una conjugación eficaz depende de la fase de crecimiento de las células receptoras, pero no se ha encontrado evidencias de esto. Otra posibilidad que barajamos es que Phr^*_{LS20} puede acumularse dentro de las células en fase estacionaria y se mantiene en niveles suficientemente altos como para inhibir Rap_{LS20} durante varias generaciones después de que las células sean diluidas. Esta posibilidad cobra fuerza al observar que los niveles de conjugación son elevados durante la fase de crecimiento exponencial cuando un cultivo de células donadoras es crecido por la noche y diluido, nuevamente crecido hasta la fase exponencial y diluido una vez más antes de comenzar el experimento.

Un resumen de los circuitos de regulación de los genes de conjugación de pLS20cat se puede observar esquemáticamente en la Figura R18. Rco_{LS20} es responsable de mantener la conjugación en estado inactivo por defecto, mediante la represión de los genes de conjugación. Rap_{LS20} puede activar la conjugación mediante la eliminación de la represión producida por Rco_{LS20}, pero sólo es capaz de hacerlo cuando su actividad no es inhibida por el péptido señal Phr^*_{LS20}. Por lo tanto, la conjugación del plásmido pLS20 está estrictamente regulada por el péptido Phr^*_{LS20} mediada por un mecanismo de quórum sensing (QS).

8- References

1. Tsutomu Watanabe, T. F. (1961) EPISOME-MEDIATED TRANSFER OF DRUG RESISTANCE IN ENTEROBACTERIACEAE III. : Transduction of Resistance Factors. *Journal of Bacteriology* **82**, 202

2. Watanabe, T. (1963) INFECTIVE HEREDITY OF MULTIPLE DRUG RESISTANCE IN. *Bacteriological reviews* **27**, 87–115

3. Pobell, S. S. (1995) *SSP-1995.pdf* (Akkermans, Antoon D.L.; Elsas, J. D. Van, ed.), Kluwer Academic Publishers, Wageningen,

4. Gotz, A., Pukall, R., Smit, E., Tietze, E., Prager, R., Tschape, H., van Elsas, J., and Smalla, K. (1996) Detection and characterization of broad-host-range plasmids in environmental bacteria by PCR. *Appl. Envir. Microbiol.* **62**, 2621–2628

5. Timms-Wilson, T. M., Bryant, K., and Bailey, M. J. (2001) Strain characterization and 16S-23S probe development for differentiating geographically dispersed isolates of the phytopathogen Ralstonia solanacearum. *Environmental Microbiology* **3**, 785–797

6. Ochman, H., Lawrence, J. G., and Groisman, E. A. (2000) Lateral gene transfer and the nature of bacterial innovation. *Nature* **405**, 299–304

7. Koonin, E. V, Makarova, K. S., and Aravind, L. (2001) H ORIZONTAL G ENE T RANSFER IN P ROKARYOTES :

8. Frost, L. S., and Koraimann, G. (2010) Regulation of bacterial conjugation: balancing opportunity with adversity. *Future microbiology* **5**, 1057–71

9. Smillie, C., Garcillán-Barcia, M. P., Francia, M. V., Rocha, E. P. C., and de la Cruz, F. (2010) Mobility of plasmids. *Microbiology and molecular biology reviews : MMBR* **74**, 434–52

10. Alvarez-Martinez, C. E., and Christie, P. J. (2009) Biological diversity of prokaryotic type IV secretion systems. *Microbiol Mol Biol Rev* **73**, 775–808

11. Fronzes, R., Christie, P. J., and Waksman, G. (2009) The structural biology of type IV secretion systems. *Nature reviews. Microbiology* **7**, 703–14

12. Grohmann, E., and Espinosa, M. (2003) Conjugative Plasmid Transfer in Gram-Positive Bacteria. **67**, 277–301

13. Koehler, T. M., and Thorne, C. B. (1987) Bacillus subtilis (natto) plasmid pLS20 mediates interspecies plasmid transfer. *Journal of bacteriology* **169**, 5271–5278

14. Itaya, M., Akaya, N. S., Atsunaga, S. M., Ujita, K. F., and Aneko, S. K. (2006) Conjugational Transfer Kinetics of pLS20 between Bacillus subtilis in Liquid Medium. **70**, 740–742

15. Garcia-Vallve, S. (2000) Horizontal Gene Transfer in Bacterial and Archaeal Complete Genomes. *Genome Research* **10**, 1719–1725

16. Frost, L. S., Leplae, R., Summers, A. O., and Toussaint, A. (2005) Mobile genetic elements: the agents of open source evolution. *Nature reviews. Microbiology* **3**, 722–32

17. Thomas, C. M., and Nielsen, K. M. (2005) Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nature reviews. Microbiology* **3**, 711–721

18. Novick, R. P., Christie, G. E., and Penadés, J. R. (2010) The phage-related chromosomal islands of Gram-positive bacteria. *Nature reviews. Microbiology* **8**, 541–551

19. Lorenz, M. G., and Wackernagel, W. (1994) Bacterial gene transfer by natural genetic transformation in the environment. *Microbiological reviews* **58**, 563–602

20. Nielsen, K. M., Johnsen, P. J., Bensasson, D., and Daffonchio, D. (2007) Release and persistence of extracellular DNA in the environment. *Environmental Biosafety Research* **6**, 37–53

21. Zafra, O., Lamprecht-Grandío, M., de Figueras, C. G., and González-Pastor, J. E. (2012) Extracellular DNA release by undomesticated Bacillus subtilis is regulated by early competence. *PloS one* **7**, e48716

22. Chen, I., and Dubnau, D. (2003) DNA transport during transformation. *Frontiers in bioscience : a journal and virtual library* **8**, s544–s556

23. Nzula, S., Vandamme, P., and Govan, J. R. (2000) Sensitivity of the Burkholderia cepacia complex and Pseudomonas aeruginosa to transducing bacteriophages. *FEMS immunology and medical microbiology* **28**, 307–312

24. Langley, R., Kenna, D. T., Vandamme, P., Ure, R., and Govan, J. R. W. (2003) Lysogeny and bacteriophage host range within the Burkholderia cepacia complex. *Journal of medical microbiology* **52**, 483–490

25. Jensen, E. C., Schrader, H. S., Rieland, B., Thompson, T. L., Lee, K. W., Nickerson, K. W., and Kokjohn, T. A. (1998) Prevalence of broad-host-range lytic bacteriophages of Sphaerotilus natans, Escherichia coli, and Pseudomonas aeruginosa. *Applied and environmental microbiology* **64**, 575–580

26. Holmfeldt, K., Middelboe, M., Nybroe, O., and Riemann, L. (2007) Large variabilities in host strain susceptibility and phage host range govern interactions between lytic marine phages and their Flavobacterium hosts. *Applied and environmental microbiology* **73**, 6730–6739

27. Montag, D., Hashemolhosseini, S., and Henning, U. (1990) Receptor-recognizing proteins of T-even type bacteriophages. The receptor-recognizing area of proteins 37 of phages T4 TuIa and TuIb. *Journal of molecular biology* **216**, 327–334

28. Hyman, P., and Abedon, S. T. (2010) Bacteriophage host range and bacterial resistance. *Advances in applied microbiology* **70**, 217–248

29. Scholl, D., Rogers, S., Adhya, S., and Merril, C. R. (2001) Bacteriophage K1-5 encodes two different tail fiber proteins, allowing it to infect and replicate on both K1 and K5 strains of Escherichia coli. *Journal of virology* **75**, 2509–2515

30. Leiman, P. G., Battisti, A. J., Bowman, V. D., Stummeyer, K., Mühlenhoff, M., Gerardy-Schahn, R., Scholl, D., and Molineux, I. J. (2007) The structures of bacteriophages K1E and K1-5 explain processive degradation of polysaccharide capsules and evolution of new host specificities. *Journal of molecular biology* **371**, 836–849

31. Transduction (2000)

32. Thierauf, A., Perez, G., and Maloy, A. S. (2009) Generalized transduction. *Methods in molecular biology (Clifton, N.J.)* **501**, 267–286

33. Burke, J., Schneider, D., and Westpheling, J. (2001) Generalized transduction in Streptomyces coelicolor. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 6289–6294

34. Casjens, S. R., Gilcrease, E. B., Winn-Stapley, D. A., Schicklmaier, P., Schmieger, H., Pedulla, M. L., Ford, M. E., Houtz, J. M., Hatfull, G. F., and Hendrix, R. W. (2005) The generalized transducing Salmonella bacteriophage ES18: complete genome sequence and DNA packaging strategy. *Journal of bacteriology* **187**, 1091–1104

35. Christie, P. J. (1997) Agrobacterium tumefaciens T-complex transport apparatus: a paradigm for a new family of multifunctional transporters in eubacteria. *Journal of bacteriology* **179**, 3085–3094

36. Levy, S. B., and Marshall, B. (2004) Antibacterial resistance worldwide: causes, challenges and responses. *Nature medicine* **10**, S122–S129

37. Dreiseikelmann, B. (1994) Translocation of DNA across bacterial membranes. *Microbiological reviews* **58**, 293–316

38. Dröge, M., Pühler, A., and Selbitschka, W. (1998) Horizontal gene transfer as a biosafety issue: a natural phenomenon of public concern. *Journal of biotechnology* **64**, 75–90

39. Gomis-Rüth, F. X., Solà, M., de la Cruz, F., and Coll, M. (2004) Coupling factors in macromolecular type-IV secretion machineries. *Current pharmaceutical design* **10**, 1551–1565

40. Fürste, J. P., Pansegrau, W., Ziegelin, G., Kröger, M., and Lanka, E. (1989) Conjugative transfer of promiscuous IncP plasmids: interaction of plasmid-encoded products with the transfer origin. *Proceedings of the National Academy of Sciences of the United States of America* **86**, 1771–1775

41. De la Cruz, F., Frost, L. S., Meyer, R. J., and Zechner, E. L. (2010) Conjugative DNA metabolism in Gram-negative bacteria. *FEMS microbiology reviews* **34**, 18–40

42. Schröder, G., and Lanka, E. (2005) The mating pair formation system of conjugative plasmids-A versatile secretion machinery for transfer of proteins and DNA. *Plasmid* **54**, 1–25

43. Dunny, G. M., and Leonard, B. a (1997) Cell-cell communication in gram-positive bacteria. *Annual review of microbiology* **51**, 527–64

44. Grohmann, E., and Espinosa, M. (2003) Conjugative Plasmid Transfer in Gram-Positive Bacteria. **67**, 277–301

45. Francia, M. V., Varsaki, A., Garcillán-Barcia, M. P., Latorre, A., Drainas, C., and de la Cruz, F. (2004) A classification scheme for mobilization regions of bacterial plasmids. *FEMS microbiology reviews* **28**, 79–100

46. Calcutt, M. J., Lewis, M. S., and Wise, K. S. (2002) Molecular Genetic Analysis of ICEF , an Integrative Conjugal Element That Is Present as a Repetitive Sequence in the Chromosome of Mycoplasma fermentans PG18. **184**, 6929–6941

47. Auchtung, J. M., Lee, C. a, Monson, R. E., Lehman, A. P., and Grossman, A. D. (2005) Regulation of a Bacillus subtilis mobile genetic element by intercellular signaling and the global DNA damage response. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 12554–9

48. Seth-smith, H., and Croucher, N. J. (2009) nEwS & An AlySIS Breaking the Ice.

49. Burrus, V., Pavlovic, G., Decaris, B., and Guédon, G. (2002) Conjugative transposons: the tip of the iceberg. *Molecular microbiology* **46**, 601–610

50. Dunny, G. M. (2007) The peptide pheromone-inducible conjugation system of Enterococcus faecalis plasmid pCF10: cell-cell signalling, gene transfer, complexity and evolution. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **362**, 1185–1193

51. Burrus, V., Marrero, J., and Waldor, M. K. (2006) The current ICE age: biology and evolution of SXT-related integrating conjugative elements. *Plasmid* **55**, 173–83

52. Lee, C. a, Thomas, J., and Grossman, A. D. (2012) The Bacillus subtilis conjugative transposon ICEBs1 mobilizes plasmids lacking dedicated mobilization functions. *Journal of bacteriology* **194**, 3165–72.

53. Grohmann, E., Muth, G., and Espinosa, M. (2003) Conjugative plasmid transfer in gram-positive bacteria. *Microbiology and molecular biology reviews : MMBR* **67**, 277–301,

54. Garcillan-Barcia, M. P., Francia, M. V., and de la Cruz, F. (2009) The diversity of conjugative relaxases and its application in plasmid classification. *FEMS Microbiology Reviews* **33**, 657–687

55. Waters, C. M., and Bassler, B. L. (2005) Quorum sensing: cell-to-cell communication in bacteria. *Annual review of cell and developmental biology* **21**, 319–346

56. Thoendel, M., and Horswill, A. R. (2010) Biosynthesis of peptide signals in gram-positive bacteria. *Advances in applied microbiology* **71**, 91–112

57. Schultz, D., Wolynes, P. G., Ben Jacob, E., and Onuchic, J. N. (2009) Deciding fate in adverse times: sporulation and competence in Bacillus subtilis. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 21027–34

58. Bose, B., Auchtung, J. M., Lee, C. a, and Grossman, A. D. (2008) A conserved anti-repressor controls horizontal gene transfer by proteolysis. *Molecular microbiology* **70**, 570–82

59. Auchtung, J. M., Lee, C. a, Garrison, K. L., and Grossman, A. D. (2007) Identification and characterization of the immunity repressor (ImmR) that controls the mobile genetic element ICEBs1 of Bacillus subtilis. *Molecular microbiology* **64**, 1515–28

60. Zhang, L., Murphy, P. J., Kerr, A., and Tate, M. E. (1993) Agrobacterium conjugation and gene regulation by N-acyl-L-homoserine lactones. *Nature* **362**, 446–448

61. Dunny, G. M., and Johnson, C. M. (2011) Regulatory circuits controlling enterococcal conjugation : lessons for functional genomics. *Current Opinion in Microbiology* **14**, 174–180

62. Clewell, D. B. (2007) Properties of Enterococcus faecalis plasmid pAD1, a member of a widely disseminated family of pheromone-responding, conjugative, virulence elements encoding cytolysin. *Plasmid* **58**, 205–27

63. Hirt, H., Wirth, R., and Muscholl, A. (1996) Comparative analysis of 18 sex pheromone plasmids from Enterococcus faecalis: detection of a new insertion element on pPD1 and implications for the evolution of this plasmid family. *Molecular & general genetics : MGG* **252**, 640–647

64. Dunny, G. M. (2013) Enterococcal sex pheromones: signaling, social behavior, and evolution. *Annual review of genetics* **47**, 457–82

65. Nakayama, J., Yoshida, K., Kobayashi, H., Isogai, A., Clewell, D. B., and Suzuki, A. (1995) Cloning and characterization of a region of Enterococcus faecalis plasmid pPD1 encoding pheromone inhibitor (ipd), pheromone sensitivity (traC), and pheromone shutdown (traB) genes. *Journal of bacteriology* **177**, 5567–5573

66. Shi, K., Brown, C. K., Gu, Z.-Y., Kozlowicz, B. K., Dunny, G. M., Ohlendorf, D. H., and Earhart, C. A. (2005) Structure of peptide sex pheromone receptor PrgX and PrgX/pheromone complexes and regulation of conjugation in Enterococcus faecalis. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 18596–18601

67. Chatterjee, A., Cook, L. C. C., Shu, C.-C., Chen, Y., Manias, D. a, Ramkrishna, D., Dunny, G. M., and Hu, W.-S. (2013) Antagonistic self-sensing and mate-sensing signaling controls antibiotic-resistance transfer. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 7086–90

68. Bensing, B. A., and Dunny, G. M. (1993) Cloning and molecular analysis of genes affecting expression of binding substance, the recipient-encoded receptor(s) mediating mating aggregate formation in Enterococcus faecalis. *Journal of bacteriology* **175**, 7421–7429

69. Kozlowicz, B. K., Shi, K., Gu, Z.-Y., Ohlendorf, D. H., Earhart, C. a, and Dunny, G. M. (2006) Molecular basis for control of conjugation by bacterial pheromone and inhibitor peptides. *Molecular microbiology* **62**, 958–969

70. Wozniak, R. A. F., and Waldor, M. K. (2010) Integrative and conjugative elements: mosaic mobile genetic elements enabling dynamic lateral gene flow. *Nature reviews. Microbiology* **8**, 552–63

71. Beaber, J. W., Hochhut, B., and Waldor, M. K. (2004) SOS response promotes horizontal dissemination of antibiotic resistance genes. *Nature* **427**, 72–4

72. Celli, J., Poyart, C., and Trieu-Cuot, P. (1997) Use of an excision reporter plasmid to study the intracellular mobility of the conjugative transposon Tn916 in gram-positive bacteria. *Microbiology* **143 (Pt 4**, 1253–1261

73. Wong, J. J. W., Lu, J., and Glover, J. N. M. (2012) Relaxosome function and conjugation regulation in F-like plasmids - a structural biology perspective. *Molecular microbiology* **85**, 602–17

74. Serna, A., Espinosa, E., Camacho, E. M., and Casadesús, J. (2010) Regulation of bacterial conjugation in microaerobiosis by host-encoded functions ArcAB and sdhABCD. *Genetics* **184**, 947–58

75. Iuchi, S., and Lin, E. C. (1991) Adaptation of Escherichia coli to respiratory conditions: regulation of gene expression. *Cell* **66**, 5–7

76. Serna, A., Espinosa, E., Camacho, E. M., Casadesu, J., and Gene, D. De (2010) Regulation of Bacterial Conjugation in Microaerobiosis by Host-Encoded Functions ArcAB and SdhABCD ´. **958**, 947–958

77. Camacho, E. M., and Casadesús, J. (2002) Conjugal transfer of the virulence plasmid of Salmonella enterica is regulated by the leucine-responsive regulatory protein and DNA adenine methylation. *Molecular microbiology* **44**, 1589–98

78. Calvo, J. M., and Matthews, R. G. (1994) The Leucine-Responsive Regulatory Protein , a Global Regulator of Metabolism in Escherichia coli. *Microbiol Mol Biol Rev* **58**

79. Horodniceanu, T., Bouanchaud, D. H., Bieth, G., and Chabbert, Y. A. (1976) R Plasmids in Streptococcus agalactiae (Group B). *Antimicrobial Agents and Chemotherapy* **10**, 795–801

80. Kurenbach, B., Kopeć, J., Mägdefrau, M., Andreas, K., Keller, W., Bohn, C., Abajy, M. Y., and Grohmann, E. (2006) The TraA relaxase autoregulates the putative type IV secretion-like system encoded by the broad-host-range Streptococcus agalactiae plasmid pIP501. *Microbiology (Reading, England)* **152**, 637–45

81. Pettis, G. S., and Cohen, S. N. (1996) Plasmid transfer and expression of the transfer (tra) gene product of plasmid pIJ101 are temporally regulated during the Streptomyces lividans life cycle. *Molecular Microbiology* **19**, 1127–1135

82. Kieser, T., Hopwood, D. A., Wright, H. M., and Thompson, C. J. (1982) pIJ101, a multi-copy broad host-range Streptomyces plasmid: functional analysis and development of DNA cloning vectors. *Molecular & general genetics : MGG* **185**, 223–228

83. Possoz, C., Ribard, C., Gagnat, J., Pernodet, J. L., and Guérineau, M. (2001) The integrative element pSAM2 from Streptomyces: kinetics and mode of conjugal transfer. *Molecular microbiology* **42**, 159–166

84. Stein, D. S., Kendall, K. J., and Cohen, S. N. (1989) Identification and analysis of transcriptional regulatory signals for the kil and kor loci of Streptomyces plasmid pIJ101. *Journal of bacteriology* **171**, 5768–5775

85. Tai, J. T., and Cohen, S. N. (1993) The active form of the KorB protein encoded by the Streptomyces plasmid pIJ101 is a processed product that binds differentially to the two promoters it regulates. *Journal of bacteriology* **175**, 6996–7005

86. Earl, A. M., Losick, R., and Kolter, R. (2008) Ecology and genomics of Bacillus subtilis. *Trends in microbiology* **16**, 269–275

87. Harwood, C. R. (1992) Bacillus subtilis and its relatives: molecular biological and industrial workhorses. *Trends in biotechnology* **10**, 247–256

88. Kunst, F., et al. (1997) The complete genome sequence of the gram-positive bacterium Bacillus subtilis. *Nature* **390**, 249–256

89. Barbe, V., et al. (2009) From a consortium sequence to a unified sequence: the Bacillus subtilis 168 reference genome a decade later. *Microbiology* **155**, 1758–1775

90. Kobayashi, K., et al. (2003) Essential Bacillus subtilis genes. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 4678–4683

91. Nicolas, P., et al. (2012) Condition-dependent transcriptome reveals high-level regulatory architecture in Bacillus subtilis. *Science (New York, N.Y.)* **335**, 1103–6

92. Buescher, J. M., et al. (2012) Global network reorganization during dynamic adaptations of Bacillus subtilis metabolism. *Science (New York, N.Y.)* **335**, 1099–103

93. Titok, M. ., Chapuis, J., Selezneva, Y. ., Lagodich, A. ., Prokulevich, V. ., Ehrlich, S. ., and Jannière, L. (2003) Bacillus subtilis soil isolates: plasmid replicon analysis and construction of a new theta-replicating vector. *Plasmid* **49**, 53–62

94. Cutting, S. M. (2011) Bacillus probiotics. *Food microbiology* **28**, 214–220

95. Tanaka, T., and Koshikawa, T. (1977) Isolation and characterization of four types of plasmids from Bacillus subtilis (natto). *Journal of bacteriology* **131**, 699–701

96. Koehler, T. M., and Thorne, C. B. (1987) Bacillus subtilis (natto) plasmid pLS20 mediates interspecies plasmid transfer. *Journal of bacteriology* **169**, 5271–5278

97. Meijer, W. J., de Boer, A. J., van Tongeren, S., Venema, G., and Bron, S. (1995) Characterization of the replication region of the Bacillus subtilis plasmid pLS20: a novel type of replicon. *Nucleic acids research* **23**, 3214–3223

98. Derman, A. I., Becker, E. C., Truong, B. D., Fujioka, A., Tucey, T. M., Erb, M. L., Patterson, P. C., and Pogliano, J. (2009) Phylogenetic analysis identifies many uncharacterized actin-like proteins (Alps) in bacteria: regulated polymerization, dynamic instability and treadmilling in Alp7A. *Molecular microbiology* **73**, 534–52

99. Derman, A. I., Nonejuie, P., Michel, B. C., Truong, B. D., Fujioka, A., Erb, M. L., and Pogliano, J. (2012) Alp7R regulates expression of the actin-like protein Alp7A in Bacillus subtilis. *Journal of bacteriology* **194**, 2715–24

100. Bauer, T., Rosch, T., Itaya, M., and Graumann, P. L. (2011) Localization Pattern of Conjugation Machinery in a Gram-Positive Bacterium. *Journal of Bacteriology* **193**, 6244–6256

101. Rösch, T. C., Golman, W., Hucklesby, L., Gonzalez-Pastor, J. E., and Graumann, P. L. (2013) The presence of conjugative plasmid pLS20 affects global transcription of its Bacillus subtilis host and results in beneficial stress resistance to cells. *Applied and environmental microbiology* **49**.

102. Sambrook, J., and Russell, D. W. (2001) *Molecular Cloning: A Laboratory Manual*

103. Schaeffer, P., Millet, J., and Aubert, J. P. (1965) Catabolic repression of bacterial sporulation. *Proceedings of the National Academy of Sciences of the United States of America* **54**, 704–11

104. Hyatt, D., Chen, G.-L., Locascio, P. F., Land, M. L., Larimer, F. W., and Hauser, L. J. (2010) Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC bioinformatics* **11**, 119

105. Zhang, X.-Z., and Zhang, Y.-H. P. (2011) Simple, fast and high-efficiency transformation system for directed evolution of cellulase in Bacillus subtilis. *Microbial biotechnology* **4**, 98–105 [

106. Parkhomchuk, D., Borodina, T., Amstislavskiy, V., Banaru, M., Hallen, L., Krobitsch, S., Lehrach, H., and Soldatov, A. (2009) Transcriptome analysis by strand-specific sequencing of complementary DNA. *Nucleic acids research* **37**, e123

107. Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**, R25

108. Ewing, B., Hillier, L., Wendl, M., and Green, P. (1998) Base-calling of automated sequencer traces usingPhred. I. Accuracy assessment. *Genome research*, 175–185

109. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) Basic local alignment search tool. *Journal of molecular biology* **215**, 403–410

110. Huang, W., Umbach, D. M., Vincent Jordan, N., Abell, A. N., Johnson, G. L., and Li, L. (2011) Efficiently identifying genome-wide changes with next-generation sequencing data. *Nucleic acids research* **39**, e130

111. Pavlidis, P., and Noble, W. S. (2003) Matrix2png: a utility for visualizing matrix data. *Bioinformatics (Oxford, England)* **19**, 295–296

112. Singh, P. K., Ballestero-Beltrán, S., Ramachandran, G., and Meijer, W. J. J. (2010) Complete nucleotide sequence and determination of the replication region of the sporulation inhibiting plasmid p576 from Bacillus pumilus NRS576. *Research in microbiology* **161**, 772–82

113. Willettslt, N., and Wilkins, B. (1984) Processing of Plasmid DNA During Bacterial Conjugation. **48**, 24–41

114. Hoa, T. T., Tortosa, P., Albano, M., and Dubnau, D. (2002) Rok (YkuW) regulates genetic competence in Bacillus subtilis by directly repressing comK. *Molecular microbiology* **43**, 15–26

115. Zavilgelsky, G. B., and Rastorguev, S. M. (2007) DNA mimicry by proteins as effective mechanism for regulation of activity of DNA-dependent enzymes. *Biochemistry. Biokhimiĭa* **72**, 913–9, 4 p. following 982

116. Nekrasov, S. V, Agafonova, O. V, Belogurova, N. G., Delver, E. P., and Belogurov, A. a (2007) Plasmid-encoded antirestriction protein ArdA can discriminate between type I methyltransferase and complete restriction-modification system. *Journal of molecular biology* **365**, 284–97

117. Turgay, K., Hahn, J., Burghoorn, J., and Dubnau, D. (1998) Competence in Bacillus subtilis is controlled by regulated proteolysis of a transcription factor. *The EMBO journal* **17**, 6730–6738

118. Hahn, J., Luttinger, A., and Dubnau, D. (1996) Regulatory inputs for the synthesis of ComK, the competence transcription factor of Bacillus subtilis. *Molecular microbiology* **21**, 763–775

119. Smits, W. K., Eschevins, C. C., Susanna, K. a, Bron, S., Kuipers, O. P., and Hamoen, L. W. (2005) Stripping Bacillus: ComK auto-stimulation is responsible for the bistable response in competence development. *Molecular microbiology* **56**, 604–14

120. Smits, W. K., and Grossman, A. D. (2010) The transcriptional regulator Rok binds A+T-rich DNA and is involved in repression of a mobile genetic element in Bacillus subtilis. *PLoS genetics* **6**, e1001207

121. D'Souza, C., Nakano, M. M., and Zuber, P. (1994) Identification of comS, a gene of the srfA operon that regulates the establishment of genetic competence in Bacillus subtilis. *Proceedings of the National Academy of Sciences of the United States of America* **91**, 9397–9401

122. Hamoen, L. W., Eshuis, H., Jongbloed, J., Venema, G., and van Sinderen, D. (1995) A small gene, designated comS, located within the coding region of the fourth amino acid-activation domain of srfA, is required for competence development in Bacillus subtilis. *Molecular microbiology* **15**, 55–63

123. Albano, M., Smits, W. K., Ho, L. T. Y., Kraigher, B., Mandic-mulec, I., Kuipers, O. P., and Dubnau, D. (2010) The Rok Protein of Bacillus subtilis Represses Genes for Cell Surface and Extracellular Functions. **187**, 2010–2019

124. Smillie, C., Garcillán-Barcia, M. P., Francia, M. V., Rocha, E. P. C., and de la Cruz, F. (2010) Mobility of plasmids. *Microbiology and molecular biology reviews : MMBR* **74**, 434–52

125. Perego, M., Hanstein, C., Welsh, K. M., Djavakhishvili, T., Glaser, P., and Hoch, J. A. (1994) Multiple protein-aspartate phosphatases provide a mechanism for the integration of diverse signals in the control of development in B. subtilis. *Cell* **79**, 1047–1055

126. Solomon, J. M., Lazazzera, B. A., and Grossman, A. D. (1996) Purification and characterization of an extracellular peptide factor that affects two different developmental pathways in Bacillus subtilis. *Genes & development*, 2013–2024

127. Perego, M., and Hoch, J. A. (1996) Cell-cell communication regulates the effects of protein aspartate phosphatases on the phosphorelay controlling development in Bacillus subtilis. **93**, 1549–1553

128. Meijer, W. J., et al. (1998) Rolling-circle plasmids from Bacillus subtilis: complete nucleotide sequences and analyses of genes of pTA1015, pTA1040, pTA1050 and pTA1060, and comparisons with related plasmids from gram-positive bacteria. *FEMS microbiology reviews* **21**, 337–68

129. Bongiorni, C., Stoessel, R., Shoemaker, D., and Perego, M. (2006) Rap Phosphatase of Virulence Plasmid pXO1 Inhibits Bacillus anthracis Sporulation †. *Journal of Bacteriology* **188**, 487–498

130. Parashar, V., Konkol, M. a, Kearns, D. B., and Neiditch, M. B. (2013) A plasmid-encoded phosphatase regulates Bacillus subtilis biofilm architecture, sporulation, and genetic competence. *Journal of bacteriology* **195**, 2437–48

131. Chen, I., Christie, P. J., and Dubnau, D. (2005) The ins and outs of DNA transfer in bacteria. *Science (New York, N.Y.)* **310**, 1456–1460

132. Duitman, E. H., Wyczawski, D., Boven, L. G., Venema, G., Kuipers, O. P., Hamoen, L. W., and Synthetases, S. (2007) Novel Methods for Genetic Transformation of Natural Bacillus subtilis Isolates Used To Study the Regulation of the Mycosubtilin and Surfactin Synthetases.Applied and Environmental Microbiology,73,11.

133. Parashar, V., Mirouze, N., Dubnau, D. A., and Neiditch, M. B. (2011) Structural Basis of Response Regulator Dephosphorylation by Rap Phosphatases. *PLoS biology* **9,** e1000589

134. Baker, M. D., and Neiditch, M. B. (2011) Structural basis of response regulator inhibition by a bacterial anti-activator protein. *PLoS biology* **9**, e1001226

135. Burbulys, D., Trach, K. A., and Hoch, J. A. (1991) Initiation of sporulation in B. subtilis is controlled by a multicomponent phosphorelay. *Cell* **64**, 545–552

136. Smits, W. K., Bongiorni, C., Veening, J., Hamoen, L. W., Kuipers, O. P., and Perego, M. (2007) Temporal separation of distinct differentiation pathways by a dual specificity Rap-Phr system in Bacillus subtilis. **65**, 103–120

137. Parashar, V., Jeffrey, P. D., and Neiditch, M. B. (2013) Conformational change-induced repeat domain expansion regulates Rap phosphatase quorum-sensing signal receptors. *PLoS biology* **11**, e1001512

138. Pottathil, M. B. A. (2003) THE EXTRACELLULAR PHR PEPTIDE-RAP PHOSPHATASE SIGNALING CIRCUIT OF BACILLUS SUBTILIS and. *Frontiers in Bioscience* **8**, 32–45

139. Pacurar, D. I., Thordal-Christensen, H., Păcurar, M. L., Pamfil, D., Botez, M. L., and Bellini, C. (2011) Agrobacterium tumefaciens: From crown gall tumors to genetic transformation. *Physiological and Molecular Plant Pathology* **76**, 76–81

140. Bose, B., and Grossman, A. D. (2011) Regulation of horizontal gene transfer in Bacillus subtilis by activation of a conserved site-specific protease. *Journal of bacteriology* **193**, 22–9

141. Strauch, M. A., and Hoch, J. A. (1993) Transition-state regulators: sentinels of Bacillus subtilis post-exponential gene expression. *Molecular microbiology* **7**, 337–342

9- Appendix

Supplemental Table 1. Characteristics of putative ORFs of pLS20cat plasmid				
ORF	**Start-stop**	**Size (aa)**	**Putative function and/or most relevant homology** **-Microorganism (amino acid identity %)** **-Putative conserved domain (PCD)**	**Putative RBS**
1	49821-49925	34	Hypothetical protein No significant homology	AAGAGG-12-ATG
2c	50193-50444	83	Hypothetical protein No significant homology	AAAGGAGT-8-ATG
3c	50446-50946	166	Hypothetical protein TROLL 208 *Bacillus phage Troll* (41%)	AAGGAGA-11-ATG
4c	51021-51263	80	Hypothetical protein No significant homology	AGGGG-10-ATG
5c	51241-51729	162	Hypothetical protein No significant homology	AAAGGAGA-7-ATG
6c	51752-51982	76	Hypothetical protein No significant homology	CGGATCG-9-GTG
7c	52811-53066	85	Hypothetical protein *Bacillus phage SP10* (44%)	AAAGGAG-6-ATG
8c	53093-53383	96	Hypothetical protein No significant homology	AAGGAGG-10-ATG
9c	53383-538389	169	Hypothetical protein *Bacillus cereus* (40%) GDA1_CD39 superfamily	GTGTG-14-ATG
10c	54714-54872	52	Hypothetical protein No significant homology	AAGGAGG-10-ATG
11	55094-55735	213	Hypothetical protein No significant homology	TGGAAAG-10-ATG
12	55849-56202	117	Hypothetical protein *Bacillus sonorensis* (38%) YolD-like protein superfamily	TGAGG-10-GTG
13	56195-56755	186	Uvr repair protein (N-terminal) B. subtilis 168 (67%)	AGGAGA-8-ATG
14	56800-57450	216	Chloramphenicol acetyltransferase (cat)	AGGAGG-8-ATG
15	57607-58479	264	Uvr repair protein (C-terminal) *B. subtilis natto*	Part of gene 13
16	58879-59036	52	Hypothetical protein No significant homology	AAGGAGG-9-GTG
17	59220-59555	111	Bacteriophage protein *B. subtilis* (64%)	AGGGAG-8-ATG
18c	59592-60209	205	Hypothetical protein *B. azotoformans* (37%)	AAAGGAGC-8-ATG
19	60623-60850	75	Hypothetical protein No significant homology	TGTAC-15-GTG
20c	60924-61550	208	VanZ- like protein *Clostridium bartlettii* (40%) VanZ superfamily	AGGGTG-5-GTG
21c	61583-61972	129	Hypothetical Protein *B. subtilis* (38%)	AAAGGAGG-8-TTG
22	62308-62580	90	Hypothetical protein *B. licheniformis* (70%)	GGGGG-11-ATG
23c	62873-63277	134	Actin-like protein Alp7R	AGGTG-7-ATG
24c	63288-64463	391	Actin-like protein Alp7A	AAGGAGA-10-ATG
25	75-1181	368	Rap (response regulator aspartate phosphatase)	GGAGG-7-ATG

			TPR superfamily domain	
26	1178-1312	44	Putative prepro-Phr signaling peptide	AACGGAGG-7-ATG
27c	1483-1968	161	Repressor protein HTH-XRE superfamily	GAGGTGG-5-GTG
28	2564-3082	172	Hypothetical protein LGRDSM20601_p0002 *Listeria grayi* (24%)	AGGGAGG-7-TTG
29	3466-4638	390	Hypothetical protein LMOh7858_pLM80_0004 *Listeria monocytogenes* str. 4b H7858 (29%)	AAAGGTGT-11-ATG
30	4705-4941	99	Pirin domain protein *Pseudomonas fulva* 12-X (32%) Tfp pilus assembly protein PilV	GCTGG-7-GTG
31	4938-5273	111	Hypothetical protein HMPREF1012_02441 *Bacillus* sp. BT1B_CT2 (70%) Prokaryotic membrane lipoprotein lipid attachment site	AAGAACG-6-TTG
32	5334-5822	162	Hypothetical protein BCAH1134_A0022 *Bacillus cereus* AH1134 (37%)	AAAGGAGA-5-ATG
33	6124-6336	70	ABC transporter permease Serratia plymuthica PRI-2C (30%)	AAATGAGGTGG-4-ATG
34	6519-8780	778	Conserved hypothetical protein *L. monocytogenes* str. 4b H7858 (53%)	GCAGCGC-8-GTG
35	9036-9398	120	Sensor histidine kinase GraS *Staphylococcus epidermidis* NIHLM020 (23%)	AAAGGGAAGG-2-ATG
36	9446-9826	126	Hypothetical protein LM5578_p34 *L. monocytogenes* 08-5578 (46%)	AAATCGAGG-7-ATG
37	9839-11047	402	Type II/IV secretion system protein Listeria monocytogenes str. 4b H7858 (51%) Flp pilus assembly protein, ATPase CpaF multidomain	AGGAGG-7-ATG
38	11049-11852	267	Hypothetical protein bcere0018_53120 *B. cereus* Rock1-15 (46%)	AGGGGG-5-GTG
39	11849-12586	245	conserved hypothetical protein *L. monocytogenes* str. 4b H7858 (42%)	AATGGAGG-7-ATG
40	12609-12971	120	Conserved hypothetical protein *L. grayi* DSM 20601 (44%) Domain of unknown function (DUF4320)	AAAGAGG-8-ATG
41	13191-13334	47	Hypothetical protein pAW63_024 *B. thuringiensis* serovar kurstaki (53%)	AGGAGA-8-ATG
42	13361-13504	47	Hypothetical protein pAW63_024 *B. thuringiensis* serovar kurstaki (53%)	AGGAGA-9-ATG
43	13560-13775	71	Hypothetical protein	AGGCGG-7-ATG
44	13806-13997	63	Hypothetical protein	AAGGAGG-8-ATG
45	14035-14265	76	Hypothetical protein TEH_20450 *Tetragenococcus halophilus* NBRC 12172 (43%)	AAGAGG-8-ATG
46	14286-15743	485	Hypothetical protein	GGGGG-7-ATG
47	15770-16861	363	Hypothetical protein	AAAGGAGG-5-TTG
48	16910-19270	786	VirD4 component of type IV secretory pathway	TGAGG-7-TTG
49	19389-22208	974	Transmembrane protein *L. monocytogenes* FSL J1-208 (42%)	GGTGG-6-ATG
50	22205-22516	103	Hypothetical protein LmonocyFSL_00185 *L. monocytogenes* FSL J1-208 (34%)	AAAGGGGG-9-ATG
51	22518-23096	192	TraE, Type IV secretion system protein *L. monocytogenes* FSL J1-208 (33%)	AGGAGG-6-ATG
52	23110-25005	631	TraE, Type IV secretion system protein *L. monocytogenes* FSL J1-208 (47%)	AGGCGG-9-ATG

53	25005-25631	208	Putative lipoprotein LpqB *L. monocytogenes* FSL J1-208 (24%)	AAGGTGG-6-ATG
54	25624-26730	368	Putative lipoprotein *L. monocytogenes* FSL J1-208 (56%) Lysozyme_Like superfamily and NLPC_P60 superfamily	AAAGGGTTG-7-ATG
55	26749-27558	269	Hypothetical protein HA1_05612 *Clostridium perfringens* F262 (32%)	AAAGGAGG-12-GTG
56	27893-28132	79	CopG-like DNA binding protein Hypothetical protein LMIV_p039 *L. monocytogenes* FSL J1-208 (32%)	AAAGGAGC-8-ATG
57	28315-28758	147	Hypothetical protein LmonocyFSL_00140 *L. monocytogenes* FSL J1-208 (37%)	TGAGG-9-ATG
58	28758-29990	410	Hypothetical protein LmonocyFSL_00135 *L. monocytogenes* FSL J1-208 (52%)	AAAGGTGA-6-ATG
59	30021-30491	156	Putative relaxase Hypothetical protein CJD_A0352 *C. perfringens* D str. JGS1721 (29%)	AAAGGAGA-8-ATG
60	30566-31945	459	Zinc beta-ribbon domain containing protein *L. monocytogenes* FSL J1-208 (42%) Superfamily :Topoisomerase-primase domain, DNA primase, catalytic core	AAAGGGGG-7-ATG
61	32052-32207	51	Hypothetical protein *B. amyloliquefaciens* LL3 (63%)	AAAGGGAGG-8-ATG
62	32208-32720	170	Toxin-antitoxin system, toxin component, MazF family *Streptococcus mitis* bv. 2 str. SK95 (37%)	AGGAAT-8-TTG
63	32938-33285	115	Hypothetical protein Bsb_33 *B. subtilis* (28%)	AAGGGAG-8-TTG
64	33969-34367	132	Rok, ComK repressor *B. subtilis* 168	AAAGGAGA-7-ATG
65	34536-34829	97	Transcriptional regulator *Desulfomonile tiedjei* DSM 6799 (29%)	AAAAGGGGT-6-ATG
66	34823-35293	156	Single-strand DNA binding protein *B. subtilis* subsp. Natto Ssb superfamily protein	GGAGGG-9-ATG
67	35318-35767	149	Hypothetical protein bthur0013_63530 *B. thuringiensis* IBL 200 (28%)	AAAGGAGA-9-ATG
68	35882-36097	71	Putative serine/threonine sodium symporter *B. subtilis* subsp. natto	AGGAGG-7-ATG
69	36219-36788	189	Recombinase/integrase *Enterococcus casseliflavus* EC30(49%)	AAGAGG-8-ATG
70	36817-37056	79	ABC-type Mn2+/Zn2+ transport systems permease component *Bifidobacterium animalis* sub (36%)	AGGGGG-5-ATG
71	37069-37731	220	Hypothetical protein	AAAGGAGA-4-ATG
72	37794-38111	105	YhgE/Pip C-terminal domain protein *Eggerthella* sp. HGA1 (23%)	AAGGGG-11-ATG
73	38388-39431	347	Hypothetical protein *Paenibacillus* sp. JDR-2 (23%)	GGAAG-9-ATG
74	39424-39669	81	rplI gene product *Candidatus Blochmannia pennsylvanicus* str. BPEN (31%)	AGGAGG-8-ATG
75c	40896-40084	270	ATP-dependent DNA ligase *B. subtilis* (74%) DNA ligase superfamily	AGGGAGG-7-TTG

76c	41920-42387	155	Hypothetical protein *B. amyloliquefaciens* (46%)	GGGAGA-8-ATG
77c	42384-43181	265	Hypothetical protein *B. cereus* (33%)	AAAGGTGA-8-ATG
78c	43186-43713	175	Nuclease protein *Caldalkalibacillus thermarum* (57%) SNase superfamily	AAAGGAGA-6-ATG
79c	43707-44000	97	Hypothetical protein No significant homology	AAAGGAGA-7-ATG
80	44099-44267	56	Hypothetical protein No significant homology	ACGAGC-9-GTG
81c	44560-44817	85	Hypothetical protein No significant homology	AAGGAGG-7-ATG
82c	44917-45759	280	Antirestriction protein *B. pseuodfirmus* (55%) Antirestriction superfamily	AAAGGAGA-7-TGT
83	45857-46027	56	Hypothetical protein No significant homology	ACGAGC-9-GTG
84	46112-46402	96	Hypothetical protein *B.subtilis subsp.* Natto (99%)	AAAGGGT-6-ATG
85	46695-46844	49	Hypothetical protein No significant homology	AAGGAGG-9-ATG
86c	47139-47333	64	Hypothetical protein *Bacillus sp*. BT1B CT2 (58%)	AAGGAGT-7-ATG
87c	47380-47511	43	Hypothetical protein No significant homology	CGGAGG-8-ATG
88c	47597-48448	283	Hypothetical protein *Bacteroides salanitronis* DSM 18170 (25%) Conserved domain of unknown function (DUF)	AAAGGAGT-9-ATG
89c	48448-48705	85	Hypothetical protein No significant homology	AAGGGG-9-GTG
90c	48798-49037	79	Hypothetical protein No significant homology	GTGTGC-13-ATG
91	49079-49249	56	Hypothetical protein No significant homology	TATGAGC-9-ATG
92	49334-49660	108	Hypothetical protein No significant homology	AAAGGGT-6-ATG

Supplemental Table 2. Strains		
Strains	**Genotype or description**	**Reference or source**
E. coli		
XL1-Blue	endA1 gyrA96(nal^R) thi-1 recA1 relA1 lac glnV44 F'[::Tn10 proAB$^+$ lacI^q Δ(lacZ)M15] hsdR17(r_{K^-} m_{K^+})	Bullock et al., 1987
BL21(DE3)	F- dcm ompT hsdS(r_{B^-} m_{B^-}) gal	Studier and Moffat, 1986
B. subtilis		
BEST40401	*hsdR hsdM leu arg* containing plasmid pLS20cat	Itaya M. et al., 2006
8G5 ComGgfp	8G5 P_{comG}-*gfp1* (kan)	Smits et al., 2005
WKS 1039	AG174 Δ*rok*$_{Bsub}$::cat(cm)	Smits and Grossman, 2010
1A976	his nprE18 aprE3 eglS Δ102 bglT/bglS ΔEV lacA::PxylA-*comK* (em)	Zhang and Zhang, 2011
8G33	8G5 *comk-lacZ(kan)*	van Sinderen et al., 1994
OKB143	*pheA1*, *srf-lacZ*	Nakano et al., 1988
168 (1A700)	*trpC2*	BGSC*
AK3	*oppA*::Tn10 (spc)	Yazgan A et al., 2001
PY79(1A747)	prototrophic attICEBs1 attSPbeta	BGSC
PKS1	*trpC2* amyE::P_{spank}-*rok*$_{LS20}$(spc)	This work
PKS3	*trpC2 thrC::P$_C$-lacZ (em)*	This work
PKS7	*trpC2 thrC*::Δ *LacZ(em)*	This work
PKS8	*trpC2 thrC::P$_C$-lacZ (em)* containing plasmid pLS20cat	This work
PKS9	*trpC2 amyE*::P_{spank}*rco*$_{LS20}$ (spc)	This work
PKS11	*trpC2* containing plasmid pLS20cat	This work
PKS14	*trpC2 amyE*::P_{spank}*rco*$_{LS20}$ (spc) containing plasmid pLS20cat	This work
PKS16	*trpC2* P_{comG}-*gfp1* (kan)	This work
PKS17	*trpC2* P_{comG}-*gfp1*(kan) containing plasmid pLS20cat	This work
PKS18	*trpC2* amyE::P_{spank}-*rok*$_{LS20}$(spc) P_{comG}-*gfp1(*kan)	This work
PKS21	*trpC2* Δ*rok*$_{Bs168}$::cat(cm)	This work
PKS22	*trpC2* amyE::P_{spank}-*rok*$_{LS20}$(spc) P_{comG}-*gfp1*(kan) Δ*rok*$_{Bs168}$::cat(cm)	This work
PKS44	*trpC2* Δ*rok*$_{Bs168}$::cat(cm), amyE::P_{xyl}-*rok*$_{LS20}$-*iyfp*	This work
PKS51	*trpC2* Δ*rok*$_{Bs168}$::cat::spec (spc)	This work
PKS53	*trpC2* Δ*rok*$_{Bs168}$::cat::spec (spc) containing plasmid pLS20cat	This work
PKS56	his nprE18 aprE3 eglS Δ102 bglT/bglS ΔEV lacA::P_{xylA}-*comK* (em) containing plasmid pLS20cat	This work
PKS57	*his npr*E18 *apr*E3 *egl*S Δ102 *bgl*T/*bgl*S ΔEV *lac*A::P_{xylA}-*comK* (em) *amyE::P$_{spank}$rco$_{LS20}$(spc)*	This work
PKS58	*his npr*E18 *apr*E3 *egl*S Δ102 *bgl*T/*bgl*S ΔEV *lac*A::P_{xylA}-*comK* (em) *amyE::P$_{spank}$rco$_{LS20}$(spc)* containing pLS20cat plasmid	This work
PKS59	*his npr*E18 *apr*E3 *egl*S Δ102 *bgl*T/*bgl*S ΔEV *lac*A::P_{xylA}-*comK* (em) *amyE::P$_{spank}$rco$_{LS20}$(spc)* containing pLS20rco plasmid	This work
PKS66	*trpC2* amyE::Pspank-*rok*$_{LS20}$(spc) *comK-lacZ*(kan)	This work
PKS67	*trpC2 comK-lacZ*(kan)	This work
PKS69	*trpC2* containing plasmid pLS20rok	This work
PKS70	*trpC2 comK-lacZ*(kan) containing plasmid pLS20cat	This work
PKS71	*trpC2 comK-lacZ*(kan) containing plasmid pLS20rok	This work
PKS72	*trpC2 rok*$_{Bs168}$::cat::spc(spc) *ComK-lacZ* (kan)	This work
PKS73	*trpC2 rok*$_{Bs168}$::cat::spc(spc) *ComK-lacZ* (kan) containing plasmid pLS20rok	This work
PKS74	*trpC2 rok*$_{Bs168}$::cat::spc(spc) *ComK-lacZ* (kan) containing plasmid pLS20rok	This work
PKS77	*his npr*E18 *apr*E3 *egl*S Δ102 *bgl*T/*bgl*S ΔEV *lac*A::P_{xylA}-*comK* (em) containing plasmid pLS20rap	This work
PKS79	*trpC2* containing plasmid pLS20rap	This work
PKS80	*trpC2 lacA*::P_{xyl}-*comK* (em)	This work
PKS81	*trpC2 lacA*::P_{xyl}-*comK* (em), *amyE*::P_{spank}-*rok*$_{LS20}$ (spc)	This work
PKS82	*trpC2 lacA*::P_{xyl}-*comK* (em), Δ*comK*::Kan, *amyE*::P_{spank}-*rok*$_{LS20}$ (spc)	This work
PKS87	*trpC2 amyE::P$_{spank}$ rap$_{LS20}$(spc)* containing plasmid pLS20rap	This work
PKS86	*trpC2 amyE*::P_{spank}rco$_{LS20}$ (spc) containing pLS20rco	This work
PKS97	*trpC2 oppA::Tn10 (spc)*	This work
PKS98	*trpC2 oppA::Tn10 (spc)*containing plasmid pLS20cat	This work
PKS101	*trpC2* amyE::P_{spank}-*rok*$_{LS20}$(spc), *srfA-lacZ* (Cm)	This work
PS110	*trpC2 amyE::*P_{spank}-*Δ(spc)*	This work

PKS113	*his npr*E18 *apr*E3 *egl*S Δ102 *bgl*T/*bgl*S ΔEV *lac*A::P_{xylA}-*comK* (em) *amy*E::P_{spank}*rco*$_{LS20}$(*spc*) containing plasmid pLS20phr	This work
PKS117	*trpC2 amy*E::P_{spank}*rco*$_{LS20}$(*spc*) containing plasmid pLS20phr	This work
PKS139	Prototrophic, attICEBs1 attSPbeta amyE::P_{hspank}*rapI* (*spc*)	This work
GR20	*trpC2 amy*E::P_{spank} *rap*$_{LS20}$ (spc)	Lab strain
GR23	*trpC2 amy*E::P_{spank} *rap*$_{LS20}$ (spc) containing plasmid pLS20cat	Lab strain

Supplemental Table 3: Plasmids		
Plasmids	**Description**	**Reference or source**
pDR110	*B. subtilis amyE* integration vector containing IPTG-inducible P_{spank} promoter	D. Rudner
pBEST501	*E. coli* vector containing neomycin resistance marker in multiple cloning site	Itaya et al., 1989
pSG4924	*B. subtilis amyE* integration vector containing P_{xyl}-promoter driving *iyfp* and designed for constructing translational fusion with *yfp*	Wu et al., 2009
pLS20cat	Native plasmid pLS20 labelled with Cm resistance cassette in unique SalI site.	Itaya et al., 2006
pDG1663	B. subtilis *thrC* integration vector containing promoter-less *lacZ* gene which is used for promoter screening	BGSC
pET28b(+)	His-tag *E. coli* cloning/expression vector	Novagen
pPSrok	rok_{LS20} gene is cloned in integration vector pDR110	This work
pLS20rok	pLS20cat derivative in which rok_{LS20} gene is replaced by Neomycin marker	This work
pPKSN7	Translational fusion of rok_{LS20} and *iyfp* gene in vector pSG4924	This work
pHis-Rok_{LS20}	rok_{LS20} gene cloned in frame behind sequence coding for His(6)	This work
pDRrco_{LS20}	rco_{LS20} gene cloned in integration vector pDR110	This work
pDRrap_{LS20}	rap_{LS20} gene cloned in integration vector pDR110	This work
pLS20xre	pLS20cat derivative in which xre_{LS20} gene is replaced by Neomycin marker	This work
pLS20rap	pLS20cat derivative in which rap_{LS20} gene is replaced by Neomycin marker	This work
pLS20phr	pLS20cat derivative in which phr_{LS20} gene is replaced by Neomycin marker	This work
pPKS26	*rapI* gene is cloned in the integration vector pDR111	This work
*, BGSC: Bacillus Genetic Stock Center, Department of Biochemistry, The Ohio State University, Columbus, OH, USA. (http://www.bgsc.org/)		

Supplemental Table 4: **Oligonucleotides**		
Name	**Sequence** (5′-3′)	**Purpose**
LSRok_FwHind LS_Rok_RevXba	cccaagcttCAAGACTGCCGATATCTTTAGAT gctctagaCTCCGCAATCCTGGTAGCAGCTTC	To amplify gene *rok*$_{LS20}$
oPKS3 oPK11	gggggTACCATGCTTACAGAAAGACAAGCGTTG ggggGAATTCTCCTGATCCGCTGCCTGAGCCGCTTCCTGACATTTC TTCGAATCTATAATAACCTCTT	To clone *rok*$_{LS20}$-*iyfp* fusion
oPKS16 oPKS17Pst	GCTTCAAAACATCGCGAAAAAACACATTCT ggggCATATGTAAAGATATCGGCAGTCTTGAAAGAAGATC	Amplifying *rok*$_{LS20}$ upstream region
oPKS18Eco oPKS19	ggggTCTAGAGGTTATTATAGATTCGAAGAAATGTAA CTGTTTTGGTTTCTCTCAATTCAAAGTAATC	Amplifying *rok*$_{LS20}$ downstream region
oGR8 oGR9	ttttCATATGCTTACAGAAAGACAAGCGTTGCAAGAT ttttCTCGAGTTACATTTCTTCGAATCTATAATAACC	To amplify *rok*$_{LS20}$ for cloning in pET28 vector
PcomK1 PcomK2	AGAATCCCCCCAATGCCTTTTTTATAGTAT GTTTTCTGACTCATATTATGGCCTCCATCC	To amplify *comK* promoter
PcomG1 PcomG2	GATCTTTCCGTTGAGAAAGATACTGGTCAA TCTCCTTTCAACGCATATTGTAGAAAAAGA	To amplify *comG* promoter
Xre20UpHind Xre20DnNheI	ccccAAGCTTTATTTTGCGAGGTGGTATAAGTG ccTAGCTAGCGTAAAAAAAGGACTGCACTTAGGCTAG	To amplify gene *rco*$_{LS20}$
Rap20UpSal Rap20DnNhe	ggGTCGACAATAGCTGGAGGGAAGTGTATGTT gGCTAGCTCATCCTAACGCTTCTGTTATTCTTTGAATTTGC	To amplify gene *rap*$_{LS20}$
oPKS20 oPKS21	TATGTAAAAAGGTCATGGCAGGCGAA ggggGAATTCAATTATTAAAAATTAGAAATGAATACATG	Amplifying *rco*$_{LS20}$ upstream region
oPKS22 oPKS23	ggggCTGCAGTATACCACCTCGCAAAATAAACCCTG TATAGGAATTCCAATCTTTCTTCGCATC	Amplifying *rco*$_{LS20}$ downstream region
oPKS38 oPKS39	AAAAACTACGTCATAATTTTTAAATTGTTC ttttCTGCAGTATTCAAAAACAAAACTAGTCC	Amplifying *rap*$_{LS20}$ upstream region
oPKS40 oPKS41	ttttGAGCTCCTAAAGAGCAAATTCAAAGAAT ATTTAATTCTTTATCAGCTTAATCACT	Amplifying *rap*$_{LS20}$ downstream region
oPKS53N oPKS54	GTTTGTCACCCAGCTGCCGTTAAGGGGTGT ttttCTGCAGAATTTTCTTCATCCTAACGCCTCC	Amplifying *phr*$_{LS20}$ upstream region
oPKS55N oPKS56	ttttGAGCTCTCACGGCTGCAGCTGAGGAGAAGATC ACGGAACTCTATATCAGCCAAGTATTCAGA	Amplifying *phr*$_{LS20}$ downstream region
oGR85 oGR86	ttttGCTAGCGTAAGGATGGGGGAATTTTCTTGCGG ttttGCATGCCTACTTAAAATCGCTGCTGCCGATAGAATCCGGCTG ATTTTCGTCAATAA	To amplify *rapI* gene

5′- overhang sequences are indicated in lower case and restriction sites are underlined